

Chapitre 1

Théorie des probabilités élémentaire

1.1 Promenades aléatoires

Ce premier chapitre offre un exposé synthétique des fondations mathématiques de la théorie des probabilités. Nous avons choisi volontairement de restreindre notre présentation à l'étude des phénomènes aléatoires ne prenant qu'un nombre fini de valeurs. Ce choix nous permet notamment de contourner la théorie abstraite de l'intégration de Lebesgue, sur laquelle la théorie générale des probabilités est édifiée. Dans ce contexte simplifié, la théorie des probabilités se résume à une analyse fonctionnelle élémentaire sur des espaces finis. A titre d'exemple, une expérience aléatoire ne prenant qu'un nombre fini de valeurs correspond tout simplement à la donnée d'un ensemble fini d'évènements munis d'une mesure représentant les de probabilités de réalisation de chacun.

Ce chapitre introductif est consacré à l'étude des divers objets mathématiques correspondant aux principales notions probabilistes, telles les notions de variables aléatoires, le conditionnement, et l'indépendance entre évènements. Ces modèles mathématiques permettent une analyse précise et rigoureuse de nombreux phénomènes aléatoires discrets. De plus, la terminologie probabiliste est en adéquation parfaite avec l'expérience. Ainsi l'étude de ces modèles permet d'approfondir conjointement l'analyse mathématique et l'intuition probabiliste de phénomènes aléatoires complexes.

Ce chapitre s'organise de la façon suivante :

La première section concerne l'étude des mesures de probabilités (sur des espaces d'évènements discrets). Nous insisterons sur les phénomènes aléatoires uniformes où chaque évènements peut se réaliser avec la même probabilité. Nous présenterons les modèles combinatoires d'urnes traditionnels.

La seconde section porte sur les notions de variables aléatoires (en abrégé **v.a.**), et la propriété de mesurabilité par rapport à une algèbre d'évènements. Ces modèles ensemblistes permettent de modéliser l'information apportée par la donnée d'une variable aléatoire.

La troisième et dernière section concerne l'une des notions les plus importantes, et la plus fructueuse de la théorie des probabilités, la notion de conditionnement. Cette notion intervient

dès que l'on étudie des phénomènes aléatoires sachant une information partielle sur le résultat de l'expérience.

1.1.1 Expériences aléatoires

Une expérience aléatoire discrète se décrit mathématiquement par la donnée d'un espace fini Ω . Les points $\omega \in \Omega$ représentent les évènements élémentaires. Ils sont parfois appelés les épreuves, les résultats, ou encore les aléas de l'expérience. Les sous ensembles A de Ω , sont appelés les évènements. **On dit que l'évènement A se produit lorsque $\omega \in A$.** Ainsi A se produit autant de fois qu'il y a d'évènements élémentaires dans A .

Une mesure de probabilité \mathbb{P} sur Ω , est une application \mathbb{P} de Ω dans $[0, 1]$, telle que $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$. La probabilité $\mathbb{P}(A)$ d'un évènement A , est définie par la formule

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) = \sum_{\omega \in \Omega} 1_A(\omega) \mathbb{P}(\omega)$$

On notera que l'on a $\mathbb{P}(\Omega) = 1$, et $\mathbb{P}(\emptyset) = 0$, avec la convention $\sum_{\emptyset} = 0$. Les ensembles Ω , et \emptyset , sont appelés respectivement, l'évènement certain, et l'évènement impossible. Le couple (Ω, \mathbb{P}) , est appelé un espace de probabilité, ou encore un espace de probabilités.

Exemple 1.1.1 *Amener un total de points au moins égal à 10, en jetant deux dés, est un évènement aléatoire A qui s'exprime dans l'espace produit*

$$\Omega = (\Omega_1 \times \Omega_2) \quad \text{avec} \quad \Omega_1 = \Omega_2 = \{1, \dots, 6\}$$

avec la formule

$$A = \{(4, 6), (6, 4), (5, 5)\}$$

Chaque aléa $\omega = (\omega_1, \omega_2)$, représente les résultats ω_1 , et ω_2 , des lancers du premier, et du deuxième dés. Pour des dés non pipés, la probabilité de réalisation d'un jet donné, disons $\omega = (1, 2)$ est de $\mathbb{P}[(1, 2)] = 1/36$. Ainsi la probabilité pour que A se réalise est donnée par

$$\mathbb{P}(A) = \mathbb{P}[(4, 6)] + \mathbb{P}[(6, 4)] + \mathbb{P}[(5, 5)] = 3/36 = 1/12$$

Si l'on s'intéresse uniquement aux résultats du premier dés, il est bien plus judicieux de les exprimer dans l'espace $\Omega_1 = \{1, \dots, 6\}$. Obtenir le chiffre 6, lors du lancer de ce dés, est un évènement aléatoire qui s'exprime dans Ω_1 , par la donnée du singleton $A_1 = \{6\} \subset \Omega_1$. La probabilité de réalisation de chaque aléa $\omega_1 \in \Omega_1$, correspondant au lancer du premier dés, est alors donnée par

$$\mathbb{P}_1(\omega_1) =_{\text{d'ef.}} \mathbb{P}[\{\omega_1\} \times \Omega_2] = \sum_{\omega_2 \in \Omega_2} \mathbb{P}[(\omega_1, \omega_2)] = 6/36 = 1/6$$

L'exemple précédent montre qu'un même évènement aléatoire peut s'exprimer dans des espaces probabilisés plus ou moins complexes. En pratique, l'on pourra mettre à profit cette souplesse, en choisissant l'espace probabilisé le plus simple rendant compte de l'expérience étudiée.

La représentation des évènements comme partie d'un ensemble Ω offre une correspondance précieuse entre les opérations logiques sur les réalisations des expériences, et les opérations d'inclusion/exclusion classiques de la théorie des ensembles : A tout évènement A , on associe son contraire $A^c = \Omega - A$ qui se réalise uniquement si A ne l'est pas. A tout couple d'évènements $A, B \subset \Omega$, l'évènement $A \cap B$ est celui qui se réalise uniquement si les évènements A et B se réalisent simultanément. De même, l'évènement $A \cup B$ se réalise, lorsque l'évènement A ou B se réalise. Lorsque l'on a $A \cap B = \emptyset$, on dit que les évènements A et B sont incompatibles.

Nous laissons le soin au lecteur de vérifier les formules suivantes :

- $A \cap B = \emptyset \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- $\forall A, B \subset \Omega \quad \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- $\forall A \subset \omega \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- Pour tout $A, B, C \subset \Omega$

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \{\mathbb{P}(A \cap B) + \mathbb{P}(B \cap C) + \mathbb{P}(A \cap C)\} + \mathbb{P}(A \cap B \cap C) \end{aligned}$$

Exercice 1.1.1 Vérifier les formules

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) & (A \cup B)^c &= A^c \cap B^c \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) & (A \cap B)^c &= A^c \cup B^c \end{aligned}$$

Exercice 1.1.2 Montrer (par récurrence) que pour toute collection d'évènements $A_1, \dots, A_n \subset \Omega$, nous avons

$$\mathbb{P}(A_1 \cup \dots \cup A_n) \leq \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$$

Exercice 1.1.3 Soit A, B un couple d'évènements. Montrer que

$$A \Delta B = [A \cap B^c] \cup [B \cap A^c] = (A - B) \cup (B - A)$$

représente l'évènement où exactement A ou B se produit. Vérifier que l'on a

$$1_{A \Delta B} = 1_A + 1_B - 21_{A \cap B}$$

En déduire par le calcul, puis schématiquement, la formule

$$\mathbb{P}(A \Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$$

Problème 1.1.1.1 (formule de Poincaré) Soit $A_1, \dots, A_n \subset \Omega$ une suite d'évènements.

1. Vérifier les formules suivantes

$$\begin{aligned} 1_{A_1 \cap A_2} &= 1_{A_1} \cap 1_{A_2} \\ 1_{A_1 \cup A_2} &= 1 - (1 - 1_{A_1})(1 - 1_{A_2}) = 1_{A_1} + 1_{A_2} - 1_{A_1 \cap A_2} \end{aligned}$$

2. Montrer que

$$1_{\cup_{p=1}^n A_p} = 1 - \prod_{p=1}^n (1 - 1_{A_p})$$

3. En utilisant la formule

$$\prod_{i=1}^n (1 + a_i) = 1 + \sum_{p=1}^n \sum_{1 \leq i_1 < \dots < i_p \leq n} a_{i_1} \dots a_{i_p}$$

qui est valable pour tout $n \geq 0$, et pour tout $(a_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, vérifier l'identité

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{p=1}^n \sum_{1 \leq i_1 < \dots < i_p \leq n} (-1)^{p-1} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) \quad (1.1)$$

4. Montrer que pour tout $A \subset \Omega$, $\omega \in \Omega$, et $u \in \mathbb{R}$, on a

$$(1 + u)^{1_A(\omega)} = 1 + u 1_A(\omega)$$

avec la convention $0^0 = 1$, lorsque $u = -1$, et $\omega \notin A$. En déduire l'identité

$$(1 + u)^{\sum_{i=1}^n 1_{A_i}} = \prod_{i=1}^n (1 + u 1_{A_i})$$

Par un raisonnement analogue à celui utilisé dans la question précédente, vérifier l'identité

$$\sum_{\omega \in \Omega} (1 + u)^{\sum_{i=1}^n 1_{A_i}(\omega)} \mathbb{P}(\omega) = 1 + \sum_{p=1}^n \sum_{1 \leq i_1 < \dots < i_p \leq n} u^p \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p})$$

5. Retrouver la formule de Poincaré (1.1), en posant $u = -1$.

1.1.2 Paysages uniformes

Lorsque tous les évènements élémentaires ω d'un espace probabilisé (Ω, \mathbb{P}) ont la même probabilité p , le calcul des probabilités des évènements composés $A \subset \Omega$ se réduit au comptage de nombre d'éléments dans A . Plus précisément, nous avons

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = |\Omega| \times p = 1 \implies p = 1/|\Omega|$$

et par conséquent, nous avons

$$\forall A \subset \Omega \quad \mathbb{P}(A) = |A|/|\Omega| \quad (1.2)$$

Dans les expressions précédents,

$$|A| = \text{Card}(A) = \sum_{\omega \in \Omega} 1_A(\omega)$$

désigne le cardinal d'un ensemble A . Les mesures de probabilité satisfaisant la condition (1.2) sont dites *uniformes* sur Ω . Dans ce contexte, le calcul des probabilités d'évènements s'accompagne d'une analyse combinatoire plus ou moins complexe de l'expérience aléatoire en question.

Modèles d'urnes

Dans l'expérience qui suit, n boules sont sélectionnées au hasard, dans une urne contenant N boules numérotées de 1 à N . Cette expérience peut être représentée par les n -uplets

$$\omega = (\omega_1, \dots, \omega_n) \in \{1, \dots, N\}^n$$

où chaque coordonnée ω_i désigne le numéro de la boule sélectionnée au n -ième tirage. Pour poursuivre notre discussion, il est essentiel de préciser les points suivants :

- Les boules sélectionnées sont elles remises dans l'urne ?
- L'ordre des sélections est il important ?

Par exemple, si les tirages sont sans remise, on ne peut avoir de répétitions d'indices dans les coordonnées des n -uplets $(\omega_1, \dots, \omega_n)$. Par exemple si $n = 2$, et $N = 4$, on ne peut avoir $(1, 1)$, pas plus que $(2, 2)$.

Dans les deux cas où les tirages sont avec, ou sans remise, il est aussi important de savoir si l'on considère l'ordre des sélections. Par exemple si $n = 2$, et $N = 4$, doit-on distinguer les résultats de l'expérience $\omega = (1, 2)$, et $\omega = (2, 1)$? En toute généralité, la réponse est oui. Car $\omega = (1, 2)$ représente le cas où la première boule sélectionnée est la boule numéro 1, la seconde la boule numéro 2 ; alors que l'aléa $\omega = (2, 1)$ représente le cas où la première boule sélectionnée est la boule numéro 2, la seconde la boule numéro 1. L'ordre n'est pas important lorsque l'on s'intéresse à des évènements de la forme

$$\begin{aligned} \langle 1, 2 \rangle &= \{(1, 2), (2, 1)\} \\ &= \text{“les boules no. 1 et 2 ont été sélectionnées au cours des 2 tirages”} \end{aligned}$$

Dans le cas de tirages avec remise, on remarquera que $\langle 1, 1 \rangle = (1, 1)$.

Pour distinguer ces deux situations, nous noterons

$$\omega = \langle \omega_1, \dots, \omega_n \rangle$$

les évènements élémentaires d'une expérience aléatoire où l'ordre n'est pas important ; et $\omega = (\omega_1, \dots, \omega_n)$ les évènements élémentaires d'une expérience aléatoire où l'ordre est important. Il est parfois commode d'appeler ces évènements, des **évènements ordonnés** $(\omega_1, \dots, \omega_n)$, et **désordonnés** $\langle \omega_1, \dots, \omega_n \rangle$.

1. **Sélection ordonnée avec remise** Dans le cas le plus simple où l'ordre des tirages est important, nous avons

$$\Omega = \{\omega : \omega = (\omega_1, \dots, \omega_n) \text{ avec } \omega_i \in \{1, \dots, N\}\} \implies |\Omega| = N^n$$

2. **Sélection ordonnée sans remise** Dans cette expérience, chaque boule sélectionnée au hasard est conservée à l'extérieur de l'urne. Dans le cas le plus simple où l'ordre est important, on a

$$\begin{aligned} \Omega &= \{\omega : \omega = (\omega_1, \dots, \omega_n) \text{ avec } \omega_i \in \{1, \dots, N\} \text{ et } \omega_i \neq \omega_j \ \forall i \neq j\} \\ \implies |\Omega| &= N(N-1)\dots(N-(n-1)) =_{\text{d\'ef.}} (N)_n =_{\text{d\'ef.}} A_N^n \end{aligned}$$

3. **Sélection désordonnée sans remise** Lorsque l'ordre est sans importance, et on a

$$\begin{aligned} \Omega &= \{\omega : \omega = \langle \omega_1, \dots, \omega_n \rangle \text{ avec } \omega_i \in \{1, \dots, N\} \text{ et } \omega_i \neq \omega_j \ \forall i \neq j\} \\ \implies |\Omega| &= \frac{N(N-1)\dots(N-(n-1))}{n!} = \frac{(N)_n}{n!} =_{\text{d\'ef.}} C_N^n \end{aligned}$$

Pour calculer ce dernier cardinal, il suffit de noter qu'à partir de chaque réalisation $\langle \omega_1, \dots, \omega_n \rangle$, on peut construire $n!$ événements ordonnés, en permutant les indices. Plus formellement, si \mathcal{G}_n désigne le groupe symétrique des permutations sur l'ensemble des indices $\{1, \dots, n\}$, les $n!$ événements ordonnés déduits de $\langle \omega_1, \dots, \omega_n \rangle$, correspondent à l'image de l'application suivante

$$\sigma \in \mathcal{G}_n \mapsto (\omega_{\sigma(1)}, \dots, \omega_{\sigma(n)})$$

4. **Sélection désordonnée avec remise** Cette situation est la plus complexe des quatre. L'ensemble Ω est ici donné par

$$\begin{aligned} \Omega &= \{\omega : \omega = \langle \omega_1, \dots, \omega_n \rangle \text{ avec } \omega_i \in \{1, \dots, N\}\} \\ \implies |\Omega| &= C_{N+(n-1)}^n = \frac{(N+(n-1))\dots(N+1)(N+0)}{n!} \end{aligned}$$

Le calcul de ce dernier cardinal s'effectue par récurrence sur le paramètre n . On note $|\Omega| = \alpha_N(n)$, le cardinal en question. Pour $n = 1$, on a clairement

$$\forall N \geq 1 \quad \alpha_N(1) = |\{\omega : \omega = \omega_1, \text{ avec } \omega_1 \in \{1, \dots, N\}\}| = N = C_N^1$$

Supposons que $\alpha_N(n) = C_{N+(n-1)}^n$, pour tout $N \geq 1$; et montrons que cette propriété est satisfaite au rang suivant.

Puisque l'ordre est sans importance, on peut convenir que les coordonnées ω_i , de chaque $(n+1)$ -uplet $\omega = [\omega_1, \dots, \omega_{n+1}]$, sont rangées par ordre croissant

$$\omega_1 \leq \omega_2 \leq \dots \leq \omega_n \leq \omega_{n+1}$$

Si $\omega_1 = 1$, alors il y a $\alpha_N(n)$ façons de choisir les autres composantes

$$1 \leq \omega_2 \leq \dots \leq \omega_n \leq \omega_{n+1}$$

dans l'ensemble $\{1, \dots, N\}$. Si maintenant $\omega_1 = 2$, alors il n'y a plus que $\alpha_{N-1}(n)$ façons de choisir les autres composantes

$$2 \leq \omega_2 \leq \dots \leq \omega_n \leq \omega_{n+1}$$

dans l'ensemble $\{2, \dots, N\}$. Plus généralement, lorsque $\omega_1 = i+1$, avec $i = 0, \dots, N-1$, alors il n'y a plus que $\alpha_{N-i}(n)$ façons de choisir les autres composantes

$$(i+1) \leq \omega_2 \leq \dots \leq \omega_n \leq \omega_{n+1}$$

dans l'ensemble $\{i+1, \dots, i+(N-i)\}$. En conséquence de cette dichotomie, et d'après notre hypothèse de récurrence, nous obtenons la formule

$$\alpha_N(n+1) = \sum_{i=0}^{N-1} \alpha_{N-i}(n) = \sum_{i=0}^{N-1} C_{(N-i)+(n-1)}^n$$

On remarque maintenant que

$$\begin{aligned} C_q^{p-1} + C_q^p &= \frac{q!}{((q-p)+1)! (p-1)!} + \frac{q!}{(q-p)! p!} \\ &= \frac{(q+1)!}{((q-p)+1)! p!} \times \left(\frac{p}{q+1} + \frac{(q-p)+1}{q+1} \right) = C_{q+1}^p \end{aligned}$$

Autrement dit, nous avons

$$C_q^{p-1} = C_{q+1}^p - C_q^p$$

En utilisant ce qui précède, on en conclut que

$$\begin{aligned} \alpha_N(n+1) &= \sum_{i=0}^{N-1} \left(C_{(N-i)+(n-1)+1}^{n+1} - C_{(N-i)+(n-1)}^{n+1} \right) \\ &= \sum_{i=0}^{N-1} \left(C_{(N+n)-i}^{n+1} - C_{(N+n)-(i+1)}^{n+1} \right) = C_{N+n}^{n+1} \end{aligned}$$

Exercices

Exercice 1.1.4 Une urne contient $N = N_1 + N_2$ boules, N_1 de couleur rouges, et N_2 de couleur noire. Lors d'une sélection de $n = N_1$ boules avec remise, quelle est la probabilité de choisir au moins une fois une boule rouge.

Exercice 1.1.5 Quelle est la probabilité pour qu'au moins deux étudiants dans un classe de n personnes aient la même date d'anniversaire ? On supposera que la date d'anniversaire est l'un des 365 jours, et chaque jour est équiprobable. Estimer cette probabilité lorsque $n = 31$.

Exercice 1.1.6 Un jeu de loterie est formé de n billets gagnants sur un total de N . On supposera, ce qui est souvent le cas, que le nombre total de billets est plus que le double du nombre de billets gagnants. Quelle est la probabilité de gagner au moins une fois, si l'on achète n billets ?

Exercice 1.1.7 Un jeu de loto est formé de $N = 49$ boules numérotées de 1 à 49. Six d'entre elles sont gagnantes, disons les 6 premières numérotées de 1 à 6. On sélectionne sans remise $n = 6$ boules dans cette urne. Quelle est la probabilité d'avoir choisi précisément ces 6 boules ?

1.1.3 Petit traité de lois**La loi binomiale**

Une pièce de monnaie est lancée n fois. On note $(\omega_1, \dots, \omega_n)$ les résultats obtenus. Nous conviendrons que $\omega_i = 0$ lorsque le résultat du i -ème jet est Pile, et $\omega_i = 1$ lorsque le résultat de ce lancer est Face. On supposera enfin que les n lancers sont indépendants, et la probabilité $p \in [0, 1]$, d'avoir face à un lancer, ne change pas dans le temps.

Cette expérience aléatoire peut aussi s'interpréter comme une succession de n épreuves aléatoires, ne prenant chacune que deux valeurs possibles. Ces phénomènes biphasés apparaissent dans de nombreuses situations pratiques : échec/succès, réalisation/ou non, d'une expérience répétée; fermeture/ouverture d'interrupteurs dans des circuits électriques, présence/absence d'une particularité,...

L'ensemble des évènements élémentaires associés à cette expérience est clairement donné par

$$\Omega = \{\omega : \omega = (\omega_1, \dots, \omega_n) \text{ avec } \omega_i \in \{0, 1\}\} = \{0, 1\}^n \quad (|\Omega| = 2^n)$$

On définit une probabilité \mathbb{P} sur Ω , en posant pour chaque $\omega = (\omega_1, \dots, \omega_n) \in \Omega$

$$\mathbb{P}(\omega_1, \dots, \omega_n) = \prod_{i=1}^n [p \mathbf{1}_1(\omega_i) + (1-p) \mathbf{1}_0(\omega_i)]$$

Cette probabilité peut aussi s'exprimer sous la forme suivante

$$\begin{aligned} \mathbb{P}(\omega_1, \dots, \omega_n) &= \prod_{i=1}^n [p^{1_{\omega_i}} (1-p)^{1_0(\omega_i)}] \\ &= \prod_{i=1}^n [p^{\omega_i} (1-p)^{1-\omega_i}] = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i} \end{aligned}$$

On s'intéresse maintenant aux événements où lors des n lancers de la pièce, le résultat "Face" apparaît exactement k fois

$$A_k = \{\omega : \omega = (\omega_1, \dots, \omega_n) \in \{0, 1\}^n \text{ avec } \sum_{i=1}^n \omega_i = k\}$$

avec $k \in \{0, \dots, n\}$. On vérifie aisément que

$$|A_k| = C_n^k \text{ et } \forall \omega \in A_k \quad \mathbb{P}(\omega) = p^k(1-p)^{n-k}$$

Par conséquent, on obtient

$$\forall k \in \{0, \dots, n\} \quad P(k) \stackrel{\text{d\'ef.}}{=} \mathbb{P}(A_k) = C_n^k p^k(1-p)^{n-k}$$

L'ensemble des probabilités $P = (P(k))_{k=0, \dots, n}$ est appelé *la loi binomiale de paramètres $(n; p)$* .

La loi multinomiale

Généralisons la loi binomiale, en considérant une succession de n épreuves aléatoires, pouvant prendre chacune r valeurs distinctes numérotées de 1 à r . On conviendra que ces n épreuves sont indépendantes, et la probabilité $p_i \in [0, 1]$, que le résultat soit $i \in \{1, \dots, r\}$ à chacune des épreuves, ne change pas dans le temps. On notera que le jeu de paramètres $(p_i)_{i=1, \dots, r} \in [0, 1]^r$ doit nécessairement être choisi de sorte que $\sum_{i=1}^r p_i = 1$.

L'espace des événements associé à cette expérience est donné par

$$\Omega = \{\omega : \omega = (\omega_1, \dots, \omega_n) \text{ avec } \omega_i \in \{1, \dots, r\}\} = \{1, \dots, r\}^n \quad (|\Omega| = r^n)$$

Pour chaque $i \in \{1, \dots, r\}$, on note

$$N_i(\omega) = \sum_{j=1}^n 1_i(\omega_j)$$

le nombre de fois où i apparaît dans la séquence $\omega = (\omega_1, \dots, \omega_n)$. On définit une probabilité \mathbb{P} sur Ω , en posant pour chaque $\omega = (\omega_1, \dots, \omega_n) \in \Omega$

$$\mathbb{P}(\omega_1, \dots, \omega_n) = \prod_{j=1}^n [p_1 1_1(\omega_j) + \dots + p_r 1_r(\omega_j)]$$

Cette probabilité peut aussi s'exprimer sous la forme produit suivante

$$\mathbb{P}(\omega_1, \dots, \omega_n) = \prod_{j=1}^n [p_1^{1_1(\omega_j)} \dots p_r^{1_r(\omega_j)}] = p_1^{N_1(\omega)} \dots p_r^{N_r(\omega)}$$

On s'intéresse maintenant aux événements où lors des n épreuves, les nombres i apparaît exactement n_i fois

$$A_{n_1, \dots, n_r} = \{(\omega_1, \dots, \omega_n) \in \{1, \dots, r\}^n : N_1(\omega) = n_1, \dots, N_r(\omega) = n_r\}$$

avec $n_i \in \{0, \dots, n\}$, et $\sum_{i=1}^r n_i = n$. On vérifie aisément que

$$|A_{n_1, \dots, n_r}| = C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n-(n_1+\dots+n_{r-1})}^{n_r} = \frac{n!}{n_1! \dots n_r!} =_{\text{déf.}} C_n^{n_1, \dots, n_r}$$

et

$$\forall \omega \in A_{n_1, \dots, n_r} \quad \mathbb{P}(\omega) = p_1^{n_1} \dots p_r^{n_r}$$

Par conséquent, pour toutes les possibilités de r -uplets $(n_i)_{i=1, \dots, r} \in \{0, \dots, n\}^r$, tel que $\sum_{i=1}^r n_i = n$, on obtient

$$P(n_1, \dots, n_r) =_{\text{déf.}} \mathbb{P}(A_{n_1, \dots, n_r}) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}$$

L'ensemble des probabilités $P(n_1, \dots, n_r)$ est appelé *la loi multinomiale de paramètres* $(n; (p_1, \dots, p_r))$.

La loi hypergéométrique

On considère une urne contenant m boules numérotées de 1 à m , avec m_i boules de couleur c_i , où $\{c_1, \dots, c_r\}$ désigne un jeu de r couleurs distinctes. Pour fixer les idées, on conviendra que les m_1 premières boules numérotées de 1 à m_1 sont de couleur c_1 , les m_2 boules suivantes, numérotées de $m_1 + 1$ à $m_1 + m_2$, sont de couleur c_2 , etc. Autrement dit, pour chaque indice de couleur $i \in \{1, \dots, r\}$, les numéros des m_i boules de couleur c_i sont données par les ensembles

$$C_i = m_1 + \dots + m_{i-1} + \{1, 2, \dots, m_i\}$$

Pour $i = 1$, on prend la convention $m_1 + \dots + m_{i-1} = \sum_{\emptyset} = 0$.

On sélectionne une série de $n (\leq m)$ boules **sans remise**, et l'on s'intéresse au résultat de l'expérience. Dans cette situation, l'espace des événements élémentaires est donné par

$$\Omega = \{\omega : \omega = (\omega_1, \dots, \omega_n) \text{ avec } \omega_i \in \{1, \dots, m\} \text{ et } \omega_i \neq \omega_j \quad \forall i \neq j\}$$

et l'on rappelle que

$$|\Omega| = (m)_n = m(m-1) \dots (m-(n-1))$$

Pour chaque $i \in \{1, \dots, r\}$, on note

$$N_i^n(\omega) = \sum_{j=1}^n 1_{C_i}(\omega_j)$$

le nombre de fois où la couleur c_i apparaît dans la séquence $\omega = (\omega_1, \dots, \omega_n)$. On notera que

$$\sum_{i=1}^r N_i^n(\omega) = \sum_{j=1}^n \left(\sum_{i=1}^r 1_{C_i}(\omega_j) \right) = n$$

On s'intéresse aux événements où lors des n tirages, chacune des couleurs c_i apparaît exactement n_i fois

$$A_{n_1, \dots, n_r} = \{(\omega_1, \dots, \omega_n) \in \Omega : N_1^n(\omega) = n_1, \dots, N_r^n(\omega) = n_r\}$$

avec $n_i \in \{0, \dots, n\}$, et $\sum_{i=1}^r n_i = n$.

Pour calculer le nombre d'évènements élémentaires associés à la réalisation de A_{n_1, \dots, n_r} , on observe tout d'abord qu'il y a

$$C_n^{n_1, \dots, n_r} = \frac{n!}{n_1! \dots n_r!}$$

possibilités de choisir n_i boules de couleurs c_i , lors des n tirages. Pour chaque répartition de ces jeux de couleurs lors des n tirages, il y a encore $(m_1)_{n_1}$ façons de choisir n_1 boules de couleur c_1 parmi m_1 , puis $(m_2)_{n_2}$ façons de choisir n_2 boules de couleur c_2 parmi m_2 , etc. Par conséquent, ces évènements composés ont pour cardinal

$$|A_{n_1, \dots, n_r}| = \frac{n!}{n_1! \dots n_r!} (m_1)_{n_1} \dots (m_r)_{n_r}$$

Il en découle que

$$\begin{aligned} P(n_1, \dots, n_r) &= \mathbb{P}(A_{n_1, \dots, n_r}) = \frac{\frac{n!}{n_1! \dots n_r!} (m_1)_{n_1} \dots (m_r)_{n_r}}{(m)_n} \\ &= \frac{\frac{(m_1)_{n_1}}{n_1!} \dots \frac{(m_r)_{n_r}}{n_r!}}{\frac{(m)_n}{n!}} = \frac{C_{m_1}^{n_1} \dots C_{m_r}^{n_r}}{C_m^n} \end{aligned}$$

Autrement dit, nous avons

$$P(n_1, \dots, n_r) = \frac{C_{m_1}^{n_1} \dots C_{m_r}^{n_r}}{C_m^n}$$

L'ensemble des probabilités $P(n_1, \dots, n_r)$, $n_i \geq 0$, $\sum_{i=1}^r n_i = n$, est appelé *loi hypergéométrique* de paramètres $(n : (m_1, \dots, m_r))$.

Interprétation dynamique : Lors du tirage sans remise, les probabilités d'obtenir certaines couleurs diminuent au fur et à mesure qu'elles sont sélectionnées. En effet, initialement, chacune des couleurs c_i a une probabilité m_i/m d'être choisie. Supposons maintenant que l'échantillon des k premières boules choisies soit formé de n_i^k boules de couleurs c_i , avec $\sum_{i=1}^r n_i^k = k$. Autrement dit, supposons que l'on ait

$$\forall i \in \{1, \dots, r\} \quad N_i^k(\omega_1, \dots, \omega_k) = n_i^k$$

Dans ce cas, l'urne ne contient plus que $(m_i - n_i^k)$ boules de couleurs c_i . Dans cette situation, la probabilité d'obtenir une boule de couleur c_i est désormais plus faible, et égale à

$$\frac{(m_i - n_i^k)}{(m - k)} = \frac{m_i}{m} \frac{(1 - n_i^k/m_i)}{\sum_{j=1}^r (1 - n_j^k/m)} \leq \frac{m_i}{m}$$

Par conséquent, la probabilité de choisir les n_1 premières boules de couleur c_1 , puis les n_2 boules

suivantes de couleur c_2 , etc, est donnée par la formule

$$\begin{aligned} & \mathbb{P}((\omega_i)_{1 \leq i \leq n_1} \in C_1, (\omega_i)_{n_1+1 \leq i \leq n_1+n_2} \in C_2, \dots) \\ &= \left(\frac{m_1}{m} \frac{m_1-1}{m-1} \dots \frac{m_1-(n_1-1)}{m-(n_1-1)} \right) \left(\frac{m_2}{(m-n_1)} \frac{m_2-1}{(m-n_1)-1} \dots \frac{m_2-(n_2-1)}{(m-n_1)-(n_2-1)} \right) \dots \\ &= \frac{(m_1)_{n_1}}{(m)_{n_1}} \frac{(m_2)_{n_2}}{(m-n_1)_{n_2}} \dots \frac{(m_r)_{n_r}}{(m-n_1-\dots-n_{r-1})_{n_r}} = \frac{(m_1)_{n_1} \dots (m_r)_{n_r}}{(m)_n} \end{aligned}$$

Bien entendu, il en est de même si les n_i boules de couleur c_i , sont choisies à des instants $t_1^i, \dots, t_{n_i}^i$

$$\mathbb{P}((\omega_{t_j^1})_{1 \leq j \leq n_1} \in C_1, (\omega_{t_j^2})_{1 \leq j \leq n_2} \in C_2, \dots) = \frac{(m_1)_{n_1} \dots (m_r)_{n_r}}{(m)_n}$$

Comme il y a $\frac{n!}{n_1! \dots n_r!}$ façon de choisir ses instants $t_1^i, \dots, t_{n_i}^i$, on retrouve le fait que

$$\mathbb{P}(A_{n_1, \dots, n_r}) = \frac{n!}{n_1! \dots n_r!} \frac{(m_1)_{n_1} \dots (m_r)_{n_r}}{(m)_n}$$

Approximation polynomiale : Supposons que l'on ait

$$m_1 \uparrow \infty, \dots, m_r \uparrow \infty \quad \text{avec} \quad \frac{m_1}{m} \rightarrow p_1, \dots, \frac{m_r}{m} \rightarrow p_r$$

Dans cette situation, il y a tellement de boules dans l'urne, que tout se passe comme dans un tirage avec remise! En effet, nous avons dans ce cas

$$\begin{aligned} & \left(\frac{m_1}{m} \frac{m_1-1}{m-1} \dots \frac{m_1-(n_1-1)}{m-(n_1-1)} \right) \left(\frac{m_2}{(m-n_1)} \frac{m_2-1}{(m-n_1)-1} \dots \frac{m_2-(n_2-1)}{(m-n_1)-(n_2-1)} \right) \dots \\ &= \left(\frac{m_1}{m} \right)^{n_1} \left(\frac{1-1/m_1}{1-1/m} \dots \frac{1-(n_1-1)/m_1}{1-(n_1-1)/m} \right) \\ & \quad \times \left(\frac{m_2}{m} \right)^{n_2} \left(\frac{1}{(1-n_1/m)} \frac{1-1/m_2}{(1-n_1/m)-1/m} \dots \frac{1-(n_2-1)/m_2}{(1-n_1/m)-(n_2-1)/m} \right) \dots \\ & \rightarrow p_1^{n_1} p_2^{n_2} \dots p_r^{n_r} \end{aligned}$$

On en conclut que la loi hypergéométrique convergence vers la loi multinomiale

$$\frac{n!}{n_1! \dots n_r!} \frac{(m_1)_{n_1} \dots (m_r)_{n_r}}{(m)_n} \xrightarrow{m_i \uparrow \infty} \frac{n!}{n_1! \dots n_r!} p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$$

1.2 Variables aléatoires

1.2.1 Définitions

La notion de variable aléatoire (en abrégé v.a.) permet une analyse approfondie et simplifiée des expériences aléatoires que nous avons rencontrées, et de bien d'autres à venir. La définition fonctionnelle suivante présente ces nouveaux objets comme une variable abstraite dont la valeur dépend du résultat ω de d'une expérience.

Définition 1.2.1 Une v.a. (discrète) X est une application d'un espace probabilisé (Ω, \mathbb{P}) dans un espace fini E . Lorsque $E \subset \mathbb{R}$, on dit que la v.a. est réelle.

Exemple 1.2.1 Dans l'exemple du lancer de n dés, la fonction identité

$$\begin{aligned} (X_1, \dots, X_n) : \quad \Omega = \{1, \dots, 6\}^n &\longrightarrow E = \{1, \dots, 6\}^n \\ \omega = (\omega_1, \dots, \omega_n) &\longrightarrow (X_1(\omega), \dots, X_n(\omega)) = (\omega_1, \dots, \omega_n) \end{aligned}$$

est une v.a. représentant le résultat du lancer. De même, les fonctions coordonnées

$$X_p : \omega = (\omega_1, \dots, \omega_n) \in \Omega \longrightarrow X_p(\omega) = \omega_p \in \{1, \dots, 6\}$$

représentent le résultat du lancer du p -ième dés. La somme, et le produit définis par

$$S_n(\omega) = \sum_{p=1}^n X_p(\omega) \quad \text{et} \quad \Pi_n(\omega) = \prod_{p=1}^n X_p(\omega)$$

sont à nouveau des v.a. (réelles).

Exemple 1.2.2 La v.a. indicatrice $Y = 1_A$ associée à un événement $A \subset \Omega$, représente la réalisation, ou non de cet événement. Notons que si $A = X^{-1}(B)$, avec $B \subset E$, la v.a. $Y = 1_{X^{-1}(B)} = 1_B \circ X$ représente la réalisation ou non de l'événement $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$.

Le lemme suivant est d'un usage assez fréquent dans l'étude de modèles probabilistes. Sa démonstration est immédiate!

Lemme 1.2.1 Si $X : \Omega \rightarrow E$ est une v.a. à valeurs dans E , et si $f : E \rightarrow F$ est une application arbitraire de E vers un espace fini auxiliaire F , alors l'application $f(X) = f \circ X$ est une v.a. à valeurs dans F .

La représentation canonique suivante offre une formulation algébrique précise de variables aléatoires. Sa démonstration ne pose aucun problème particulier.

Proposition 1.2.1 Soit $X(\Omega) = \{x_1, \dots, x_d\} \subset E$, l'ensemble des réalisations possibles d'une v.a. $X : \Omega \rightarrow E$. Dans cette situation, la famille d'ensembles $A_i = X^{-1}(\{x_i\})$, $1 \leq i \leq d$, forme une partition de Ω , et on a la représentation

$$X = \sum_{i=1}^d x_i 1_{A_i}$$

Exercice 1.2.1 Soit $X : \Omega \rightarrow E$ une v.a., et $f : E \rightarrow F$ une fonction de E dans un ensemble fini F . Si $X(\Omega) = \{x_1, \dots, x_d\} \subset E$, l'ensemble des réalisations possibles de v.a. X , montrer que la v.a. $f(X)$ peut s'écrire sous la forme suivante

$$f(X) = \sum_{i=1}^d f(x_i) 1_{X^{-1}(\{x_i\})}$$

Exercice 1.2.2 Soit $(A_i)_{i \leq d}$ une partition de l'ensemble Ω . Montrer que pour toute collection de nombres réels $a = (a_i)_{i \leq d}$, les fonctions

$$X^a = \sum_{i=1}^d a_i 1_{A_i}$$

sont des v.a. réelles. Vérifier que toutes les v.a. réelles sont de cette forme.

Exercice 1.2.3 Vérifier que pour tout couple d'ensembles A, B , on a les décompositions d'indicatrices suivantes

$$\begin{aligned} 1_{A \cap B} &= 1_A 1_B \\ 1_{A \cup B} &= 1_A + 1_B - 1_A 1_B = 1_A \vee 1_B \\ 1_{A - B} &= 1_A (1 - 1_B) = 1_A - 1_A 1_B \\ 1_{A \Delta B} &= (1_A - 1_B)^2 = |1_A - 1_B| \end{aligned}$$

1.2.2 Adaptation et mesurabilité

Pour toute v.a. X à valeurs dans un espace fini E , l'ensemble des événements

$$\sigma(X) = X^{-1}(\mathcal{P}(E)) = \{X^{-1}(A) : A \subset E\}$$

forme une algèbre de parties de l'ensemble Ω . Plus précisément, on vérifie que $\sigma(X)$ hérite de la structure d'algèbre sur l'ensemble des parties de E , en remarquant que

$$X^{-1}(\emptyset) = \emptyset \quad \text{et} \quad X^{-1}(E) = \Omega$$

De plus, pour tout couple de sous ensembles $A, B \subset E$, on a

$$X^{-1}(A) \cup X^{-1}(B) = X^{-1}(A \cup B) \quad \text{et} \quad X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)$$

Chaque sous ensemble

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

est formé des évènements conduisant la variable X dans l'ensemble A . Autrement dit $X^{-1}(A)$ contient tous les aléas où l'évènement $\{X \in A\}$ se réalise. Les évènements composés $X^{-1}(A) \cup X^{-1}(B) = X^{-1}(A \cup B)$ correspondent aux évènements où X prend ses valeurs dans A ou dans B . De même, évènements composés $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)$ correspondent aux évènements où X prend ses valeurs dans A et dans B .

Définition 1.2.2 Soit X une v.a. à valeurs dans un espace fini E , et définie sur un espace probabilisé (Ω, \mathbb{P}) . L'ensemble des évènements

$$\sigma(X) = X^{-1}(\mathcal{P}(E)) = \{X^{-1}(A) : A \subset E\}$$

est appelé l'algèbre sur E engendrée par la v.a. X . Ces algèbres ensemblistes représentent tous les évènements aléatoires que l'on peut décrire par des phrases grammaticales logiques.

On notera que l'algèbre $\sigma(X)$ engendrée par une v.a. X , coïncide avec la plus petite sous algèbre de $\mathcal{P}(\Omega)$ contenant les évènements

$$X^{-1}(x) \subset \Omega \quad \text{avec } x \in E$$

Pour vérifier cette assertion, on remarquera que chacun des évènements $X^{-1}(A)$, avec $A \subset E$, s'exprime sous la forme

$$X^{-1}(A) = \cup_{x \in A} X^{-1}(\{x\})$$

Par conséquent, pour toute sous algèbre $\mathcal{F} \subset \mathcal{P}(\Omega)$ contenant les évènements $(X^{-1}(x))_{x \in E}$, et d'après les propriétés de stabilité par réunion, on obtient l'inclusion

$$\sigma(X) \subset \mathcal{F}$$

L'observation précédente souligne le fait que $\sigma(X)$ correspond à l'ensemble de tous les évènements de Ω que l'on peut construire et décrire "en observant" les valeurs prises par la v.a. X . A titre d'exemple, l'ensemble

$$X^{-1}(\{x_1, x_2\}) = \{\omega \in \Omega : X(\omega) = x_1 \quad \text{ou} \quad X(\omega) = x_2\}$$

correspond à l'évènement où la v.a. X prend soit la valeur x_1 , soit la valeur x_2 . Comme nous l'avons vu dans la proposition 1.2.1, la v.a. X s'exprime alors en terme d'évènements de $\sigma(X)$ par la formule

$$X = \sum_{x \in E} x 1_{X^{-1}(\{x\})}$$

Toute algèbre $\mathcal{F} \supset \sigma(X)$ plus fine que $\sigma(X)$, contient plus d'information que nécessaire pour décrire X . En effet, supposons que

$$\mathcal{F} = \sigma(X, Y) \quad (\supset \sigma(X))$$

soit l'algèbre engendrée par un couple de v.a. (X, Y) à valeurs dans un espace produit $(E \times F)$, et définies sur un même espace de probabilités (Ω, \mathbb{P}) . Dans cette situation, nous avons la décomposition

$$X^{-1}(\{x\}) = \cup_{y \in F} (X, Y)^{-1}(\{(x, y)\})$$

Autrement dit, l'évènement $X^{-1}(\{x\})$ ne permet pas de distinguer les valeurs prises par la v.a. Y . En ce sens les évènements $X^{-1}(\{x\})$ sont plus grossiers que les évènements $(X, Y)^{-1}(\{(x, y)\})$. Pour conclure, on remarquera comme précédemment que la v.a. (X, Y) s'exprime sous la forme

$$(X, Y) = \sum_{(x, y) \in (E \times F)} (x, y) 1_{(X, Y)^{-1}(\{(x, y)\})}$$

alors que l'on a

$$\begin{aligned} X &= \sum_{(x, y) \in (E \times F)} x 1_{(X, Y)^{-1}(\{(x, y)\})} \\ &= \sum_{x \in E} x \left[\sum_{y \in F} 1_{(X, Y)^{-1}(\{(x, y)\})} \right] = \sum_{x \in E} x 1_{X^{-1}(\{x\})} \end{aligned}$$

Définition 1.2.3 Soit X une v.a. à valeurs dans un espace fini E , et définie sur un espace probabilisé (Ω, \mathbb{P}) . On dit que X est mesurable par rapport, ou adaptée à une algèbre d'évènements $\mathcal{F} \subset \mathcal{P}(\Omega)$, et on note $X \underline{\subseteq} \mathcal{F}$, lorsque l'on a

$$\forall A \subset E \quad X^{-1}(A) \subset \mathcal{F}$$

Exemple 1.2.3 La fonction identité

$$\begin{aligned} (X_1, X_2) : \Omega = \{1, \dots, 6\}^2 &\longrightarrow E = \{1, \dots, 6\}^2 \\ \omega = (\omega_1, \omega_2) &\longrightarrow (X_1(\omega), X_2(\omega)) = (\omega_1, \omega_2) \end{aligned}$$

est une v.a. représentant le résultat du lancer de deux dés. Par construction, nous avons

$$\sigma(X_1, X_2) = \mathcal{F}_2 =_{\text{déf.}} \mathcal{P}(\{1, \dots, 6\}^2)$$

On notera que

$$\forall i \in \{1, \dots, 6\} \quad X_1^{-1}(\{i\}) = \{i\} \times \{1, \dots, 6\}$$

L'algèbre engendrée par la v.a. représentant le lancer du premier dés est alors clairement donnée par

$$\begin{aligned} \mathcal{F}_1 &=_{\text{déf.}} \sigma(X_1) \\ &= \mathcal{P}(\{1, \dots, 6\}) \times \{1, \dots, 6\} = \{A \times \{1, \dots, 6\} : A \subset \{1, \dots, 6\}\} \\ &\subset \sigma(X_1, X_2) = \mathcal{P}(\{1, \dots, 6\}^2) \end{aligned}$$

avec la convention $\emptyset \times \{1, \dots, 6\} = \emptyset$. Chaque élément de l'algèbre $\sigma(X_1)$

$$\begin{aligned} B &= A \times \{1, \dots, 6\} \\ &= \{\omega \in \Omega : X_1(\omega) \in A\} \quad \text{avec } A \subset \{1, \dots, 6\} \end{aligned}$$

correspond à l'évènement "le lancer du premier dès est à valeur dans A ". On notera dans ce cas que

$$X_1 \subseteq \mathcal{F}_1 \subset \mathcal{F}_2$$

On a aussi les représentations

$$(X_1, X_2) = \sum_{i,j=1}^6 (i, j) 1_{(X_1, X_2)^{-1}(\{(i,j)\})} = \sum_{i,j=1}^6 (i, j) 1_{\{(i,j)\}}$$

et

$$X_1 = \sum_{i=1}^6 i 1_{X_1^{-1}(\{i\})} = \sum_{i=1}^6 i 1_{\{i\} \times \{1, \dots, 6\}}$$

Chaque aléa donné ω , appartient à l'un des ensembles $(\{i\} \times \{1, \dots, 6\})$, avec $i \in \{1, \dots, 6\}$. Par exemple l'aléa $\omega = (2, 5)$ représentant la situation où les lancers des dès sont 2, puis 5, appartient à l'ensemble $(\{2\} \times \{1, \dots, 6\})$. Il en est de même des aléas $\omega = (2, 1)$, $\omega = (2, 2)$, $\omega = (2, 3), \dots$, et $\omega = (2, 6)$. Dans tous ces cas, nous avons

$$1_{\{2\} \times \{1, \dots, 6\}}(\omega) = 1 \quad \text{et} \quad \forall i \neq 2 \quad 1_{\{i\} \times \{1, \dots, 6\}}(\omega) = 0$$

Autrement dit, pour tous les aléas de la forme $\omega = (2, j)$, l'évènement aléatoire $(\{2\} \times \{1, \dots, 6\}) \in \sigma(X_1)$ se réalise, et on a

$$\omega = (2, j) \implies X_1(\omega) = \sum_{i=1}^6 i 1_{\{i\} \times \{1, \dots, 6\}}(\omega) = 2$$

D'après les discussions précédentes, une v.a. donnée Y à valeurs dans un espace fini E , s'exprime de façon naturelle en terme des évènements $Y^{-1}(\{y\}) \in \sigma(Y)$, $y \in E$, par la formule

$$Y = \sum_{y \in E} y 1_{Y^{-1}(\{y\})}$$

Ces représentations événementielles peuvent clairement s'étendre à tout algèbre \mathcal{F} plus fine que $\sigma(Y)$, en ce sens où $\sigma(Y) \subset \mathcal{F}$. Par conséquent, **une algèbre contenant l'algèbre engendrée par une v.a. Y contient toute l'information relative à la réalisation de cette variable.** Il en est intuitivement de même pour toute v.a. auxiliaire X , adaptée ou mesurable par rapport à $\sigma(Y)$. La proposition suivante montre que l'adaptation d'une v.a. à une algèbre donnée est une propriété fonctionnelle très forte.

Proposition 1.2.2 Pour tout un couple de v.a. (X, Y) à valeurs dans un espace produit fini $(E \times F)$, et défini sur un même espace probabilisé (Ω, \mathbb{P}) , on a l'équivalence suivante

$$X \subseteq \sigma(Y) \iff \exists h : F \rightarrow E \quad X = h(Y)$$

Preuve:

La condition suffisante est immédiate. En effet, il suffit de noter que

$$\forall B \subset E \quad X^{-1}(B) = h(Y)^{-1}(B) = Y^{-1}(h^{-1}(B)) \in \sigma(Y)$$

Pour vérifier que toute v.a. $X \subseteq \sigma(Y)$ est nécessairement de la forme $X = h(Y)$, avec $h : F \rightarrow E$, on observe que

$$X^{-1}(\{x\}) \in \sigma(Y) \iff \exists A_x \subset F \quad X^{-1}(\{x\}) = Y^{-1}(A_x)$$

Il reste alors à noter que $1_{X^{-1}(\{x\})} = 1_{Y^{-1}(A_x)}$. En effet, en utilisant la décomposition

$$X = \sum_{x \in E} x 1_{X^{-1}(\{x\})}$$

on obtient bien $X = h(Y)$, avec la fonction $h(y) = \sum_{x \in E} x 1_{A_x}$. ■

Exemple 1.2.4 La fonction identité

$$\begin{aligned} X : \quad \Omega = \{1, \dots, 6\} &\longrightarrow E = \{1, \dots, 6\} \\ \omega &\longrightarrow X(\omega) = \omega \end{aligned}$$

est une v.a. représentant le résultat du lancer d'un dé. L'algèbre engendrée par cette v.a. coïncide clairement avec l'algèbre discrète sur Ω

$$\sigma(X) = \mathcal{P}(\Omega)$$

L'application indicatrice

$$Y = 1_{\{2,4,6\}} \circ X$$

est une v.a. représentant la parité ou non du résultat du lancer. L'algèbre engendrée par cette v.a. est donnée par

$$\sigma(Y) = \{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\} \subset \sigma(X)$$

Exercice 1.2.4 Soit $\mathcal{D} = (D_i)_{i \in I}$ une partition d'un ensemble fini Ω , muni de l'algèbre discrète $\mathcal{P}(\Omega)$. On note $a(\mathcal{D})$ la plus petite algèbre contenant les événements $(D_i)_{i \in I}$.

1. Vérifier que pour toute algèbre $b(\mathcal{D})$ contenant \mathcal{D} , on a

$$b(\mathcal{D}) \subset a(\mathcal{D}) \implies a(\mathcal{D}) = b(\mathcal{D})$$

2. On considère les v.a. indicatrices

$$\forall i \in I \quad X_i = 1_{D_i}$$

Montrer que

$$\sigma(X_i, i \in I) = \sigma(\mathcal{D})$$

L'algèbre $\sigma(\mathcal{D})$ est souvent notée abusivement $\sigma(\mathcal{D})$.

3. Montrer que pour toute fonction $f : \{0, 1\}^I \rightarrow \mathbb{R}$ nous avons

$$f((X_i)_{i \in I}) \in \sigma(\mathcal{D})$$

Vérifier que toutes les v.a. $Z \in \sigma(\mathcal{D})$ sont nécessairement de cette forme.

Exercice 1.2.5 Soit \mathcal{F} une algèbre de parties sur un ensemble fini Ω .

1. Montrer qu'il existe une partition \mathcal{D} de Ω telle que

$$\mathcal{F} = \sigma(\mathcal{D})$$

2. En déduire qu'il existe une collection finie de v.a. $(X_i)_{i \in I}$ à valeurs réelles sur Ω , telles que

$$\mathcal{F} = \sigma(X_i, i \in I)$$

1.2.3 Lois et espérances

Soit $X : \Omega \rightarrow E$ une v.a. à valeurs dans un ensemble fini E , et définie sur un espace probabilisé (Ω, \mathbb{P}) . Pour tout $x \in X(\Omega)$, le sous ensemble $X^{-1}(\{x\}) \subset \Omega$ représente l'évènement sur lequel la v.a. X prend la valeur x . La probabilité de ces évènements est donnée par la fonction

$$\mathbb{P}^X : x \in E \mapsto \mathbb{P}^X(x) = \mathbb{P}(X^{-1}(\{x\})) \in [0, 1]$$

Plus généralement, la probabilité pour que la v.a. X prenne ses valeurs dans $B \subset E$, est donnée par

$$\forall B \subset E \quad \mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B))$$

La mesure de probabilité \mathbb{P}^X sur E ainsi définie est appelée *la loi ou la distribution de la v.a. X* .

On note que

$$\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\cup_{x \in B} X^{-1}(\{x\})) = \sum_{x \in B} \mathbb{P}(X^{-1}(\{x\}))$$

On obtient ainsi une formule permettant de calculer directement la loi de chaque évènement $A = \{X \in B\} \subset \Omega$, sans passer par l'espace probabilisé (Ω, \mathbb{P})

$$\mathbb{P}^X(B) = \sum_{x \in B} \mathbb{P}^X(x)$$

Définition 1.2.4 L'espérance, ou la moyenne $\mathbb{E}(X)$ d'une v.a. réelle définie sur un espace de probabilités (Ω, \mathbb{P}) , est la quantité donnée par

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) = \sum_{x \in X(\Omega)} x \mathbb{P}^X(x)$$

Pour vérifier la seconde formulation, on note simplement que

$$\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) = \sum_{x \in X(\Omega)} x \sum_{\omega: X(\omega)=x} \mathbb{P}(\omega) = \sum_{x \in X(\Omega)} x \mathbb{P}^X(x)$$

Si $X = (X_1, X_2)$ désigne une v.a. à valeurs dans un espace produit fini $(E_1 \times E_2)$, nous avons pour toute fonction $f : E_1 \times E_2 \rightarrow \mathbb{R}$

$$\begin{aligned} \mathbb{E}(f(X_1, X_2)) &= \sum_{\omega \in \Omega} f((X_1, X_2)(\omega)) \mathbb{P}(\omega) \\ &= \sum_{(x_1, x_2) \in (E_1 \times E_2)} f(x_1, x_2) \left[\sum_{\omega: (X_1, X_2)(\omega)=(x_1, x_2)} \mathbb{P}(\omega) \right] \\ &= \sum_{(x_1, x_2) \in (E_1 \times E_2)} f(x_1, x_2) \mathbb{P}^{(X_1, X_2)}(x_1, x_2) \end{aligned}$$

Lorsque les v.a. sont réelles, on obtient clairement la propriété de linéarité

$$\forall (a_1, a_2) \in \mathbb{R}^2 \quad \mathbb{E}(a_1 X_1 + a_2 X_2) = a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2)$$

Exemple 1.2.5 Soit $\Omega = \{0, 1\}$, $\mathbb{P}(1) = 1 - \mathbb{P}(0) = p \in [0, 1]$, l'espace probabilisé associé à un jeu de pile ou face, avec probabilité de succès p . On note X la variable aléatoire canonique

$$X : \omega \in \Omega = \{0, 1\} \longrightarrow X(\omega) = \omega \in E = \{0, 1\}$$

On a ainsi de façon immédiate

$$\forall x \in E \quad \mathbb{P}^X(x) = p^x (1-p)^{1-x}$$

On note aussi que

$$\mathbb{E}(X) = 0 \times (1-p) + 1 \times p = p$$

Exemple 1.2.6 On considère l'espace $\Omega = \{0, 1\}^n$, muni de la probabilité

$$\mathbb{P}(\omega_1, \dots, \omega_n) = \prod_{i=1}^n [p \mathbf{1}_1(\omega_i) + (1-p) \mathbf{1}_0(\omega_i)]$$

Cet espace probabilisé (Ω, \mathbb{P}) correspond à une succession de n lancers d'une pièce de monnaie, ayant une probabilité de succès p . On introduit la séquence de v.a.

$$X_i : \omega = (\omega_i)_{1 \leq i \leq n} \in \Omega = \{0, 1\}^n \longrightarrow X_i(\omega) = \omega_i \in E = \{0, 1\}$$

Ces n v.a. correspondent aux résultats des i -ème lancers. La loi du vecteur aléatoire identité

$$X = (X_1, \dots, X_n) : \omega \in \Omega = \{0, 1\}^n \longrightarrow X(\omega) = \omega \in E^n = \{0, 1\}^n$$

est donnée pour tout $x = (x_i)_{1 \leq i \leq n} \in E^n$, par la formule produit

$$\mathbb{P}^X(x) = \mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = \prod_{i=1}^n \mathbb{P}^{X_i}(x_i)$$

Exemple 1.2.7 Soit $\Omega = \{-1, 1\}$, $\mathbb{P}(1) = 1 - \mathbb{P}(-1) = p \in [0, 1]$, l'espace probabilisé associé à une v.a. de Bernoulli à valeurs dans $\{-1, 1\}$. On note X la variable aléatoire canonique

$$X : \omega \in \Omega = \{-1, 1\} \longrightarrow X(\omega) = \omega \in E = \{-1, 1\}$$

On a clairement

$$\forall x \in E \quad \mathbb{P}^X(x) = p^{\frac{x+1}{2}} (1-p)^{-\frac{x-1}{2}}$$

On vérifie enfin que

$$\mathbb{E}(X) = (-1) \times (1-p) + 1 \times p = 2p - 1$$

Exemple 1.2.8 On considère l'espace $\Omega = \{-1, 1\}^n$, muni de la probabilité

$$\begin{aligned} \mathbb{P}(\omega) &= \mathbb{P}(\omega_1, \dots, \omega_n) \\ &= \prod_{i=1}^n [p \mathbf{1}_1(\omega_i) + (1-p) \mathbf{1}_{-1}(\omega_i)] \\ &= \prod_{i=1}^n p^{\frac{\omega_i+1}{2}} (1-p)^{\frac{1-\omega_i}{2}} = p^{\frac{n+\sum_{i=1}^n \omega_i}{2}} (1-p)^{\frac{n-\sum_{i=1}^n \omega_i}{2}} \end{aligned}$$

Autrement dit, on a aussi

$$\mathbb{P}(\omega) = p^{\frac{n+\sum_{i=1}^n \omega_i}{2}} (1-p)^{n-\frac{n+\sum_{i=1}^n \omega_i}{2}} = p^{N(\omega)} (1-p)^{n-N(\omega)}$$

avec

$$N(\omega) = \sum_{i=1}^n \mathbf{1}_1(\omega_i) = \frac{n + \sum_{i=1}^n \omega_i}{2}$$

On remarquera que

$$\begin{aligned} n + \sum_{i=1}^n \omega_i &= \sum_{i=1}^n (\omega_i + 1) \\ &= \sum_{i=1}^n (\mathbf{1}_1(\omega_i) + 1) + \sum_{i=1}^n (\mathbf{1}_{-1}(\omega_i) + 1) \\ &= \sum_{i=1}^n (\mathbf{1}_1(\omega_i) + 1) + 0 = 2 \sum_{i=1}^n \mathbf{1}_1(\omega_i) \end{aligned}$$

Cet espace probabilisé (Ω, \mathbb{P}) modélise les n déplacements aléatoires vers le haut, ou vers le bas, d'une particule évoluant sur \mathbb{Z} . Ce modèle probabiliste peut aussi représenter les évolutions à la hausse, ou à la baisse, du cours d'un actif financier ; les successions de pertes ou gains lors d'un processus de jeu d'argent,...

On introduit la séquence de v.a.

$$X_i : \omega = (\omega_i)_{1 \leq i \leq n} \in \Omega = \{-1, 1\}^n \longrightarrow X_i(\omega) = \omega_i \in E = \{-1, 1\}$$

Ces n v.a. correspondent aux i -ème déplacements de la particule. La loi du vecteur aléatoire identité

$$X = (X_1, \dots, X_n) : \omega \in \Omega = \{-1, 1\}^n \longrightarrow X(\omega) = \omega \in E^n = \{-1, 1\}^n$$

est ainsi donnée pour tout $x = (x_i)_{1 \leq i \leq n} \in E^n$, par la formule produit

$$\mathbb{P}^X(x) = \mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n p^{1_{\{x_i=1\}}} (1-p)^{1_{\{x_i=-1\}}} = \prod_{i=1}^n \mathbb{P}^{X_i}(x_i)$$

ou de façon équivalente, en terme du nombre de montées

$$\mathbb{P}^X(x) = p^{\frac{n+\sum_{i=1}^n x_i}{2}} (1-p)^{n-\frac{n+\sum_{i=1}^n x_i}{2}}$$

On notera enfin que

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = (2p-1)n$$

Proposition 1.2.3 Soit $X : \Omega \rightarrow E$ une v.a. définie sur un espace probabilisé (Ω, \mathbb{P}) à valeurs dans un espace fini E , et soit $f : E \rightarrow F$ une application arbitraire de E dans un espace fini auxiliaire F . Alors, la fonction composée

$$Y = f(X) = f \circ X : \omega \in \Omega \longrightarrow Y(\omega) = f(X(\omega)) \in F$$

est une v.a. définie sur le même espace probabilisé (Ω, \mathbb{P}) , et distribuée sur F selon la loi

$$\forall y \in Y \quad \mathbb{P}^Y(y) = \mathbb{P}^X(f^{-1}(\{y\})) = \sum_{x : f(x)=y} \mathbb{P}^X(x)$$

Corollaire 1.2.1 Soit (X, Y) un couple de v.a. sur un espace probabilisé (Ω, \mathbb{P}) , distribué sur un espace fini produit $(E \times F)$ selon une loi $\mathbb{P}^{(X, Y)}$. Alors, les lois respectives \mathbb{P}^X , et \mathbb{P}^Y , des v.a. individuelles X , et Y , sont données par

$$\mathbb{P}^X(x) = \sum_{y \in F} \mathbb{P}^{(X, Y)}(x, y) \quad \text{et} \quad \mathbb{P}^Y(y) = \sum_{x \in E} \mathbb{P}^{(X, Y)}(x, y)$$

Les formules précédentes s'étendent clairement au cas de plusieurs v.a. en "sommant les v.a. indésirables". Ainsi, la loi d'un vecteur (X_1, \dots, X_n) se déduit d'un autre vecteur (X_1, \dots, X_{n+m}) par la formule

$$\begin{aligned} & \mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \sum_{x_{n+1}, \dots, x_{n+m}} \mathbb{P}^{((X_1, \dots, X_n), (X_{n+1}, \dots, X_{n+m}))}((x_1, \dots, x_n), (x_{n+1}, \dots, x_{n+m})) \end{aligned}$$

Exercices

Les exercices suivants concernent de modèles d'urnes contenant m boules numérotées de 1 à m . On sélectionne successivement, et avec remise, n boules, et on note X_1, X_2, \dots, X_n les numéros des boules obtenues à la 1^{ère}, 2^{ème}, ..., et $n^{\text{ième}}$ étape.

Exercice 1.2.6 Vérifier que le n -uplet (X_1, X_2, \dots, X_n) est une v.a. pouvant être réalisée sur l'espace produit introduit à la page 12. Montrer que pour tout indice $1 \leq p \leq n$, et pour tout n -uplet $(x_1, \dots, x_n) \in E^n = \{1, \dots, m\}^n$, on a

$$\mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = 1/m^n \quad \text{et} \quad \mathbb{P}^{X_p}(x_p) = 1/m$$

Vérifier que

$$\mathbb{P}(\forall i \neq j \quad X_i \neq X_j) = (m)_n/m^n$$

Exercice 1.2.7 Supposons que parmi les m boules, m_1 sont noires et m_2 sont rouges, avec $m = m_1 + m_2$. On note respectivement B_1 , et B_2 , l'ensemble des numéros des boules noires, et des boules rouges. On considère la collection de v.a.

$$Y_i = 1_{B_1}(X_i)$$

Vérifier que pour tout $1 \leq i \leq n$, et tout n -uplet $(y_1, \dots, y_n) \in \{0, 1\}^n$, on a

$$\mathbb{P}^{Y_i}(y_i) = \left(\frac{m_1}{m}\right)^{y_i} \left(\frac{m_2}{m}\right)^{1-y_i}$$

et

$$\mathbb{P}^{(Y_1, \dots, Y_n)}(y_1, \dots, y_n) = \left(\frac{m_1}{m}\right)^{\sum_{i=1}^n y_i} \left(\frac{m_2}{m}\right)^{n - \sum_{i=1}^n y_i}$$

On pose $S_n = \sum_{i=1}^n Y_i$. Montrer que l'on a pour tout $k \in \{0, \dots, n\}$

$$\mathbb{P}^{S_n}(k) = C_n^k \left(\frac{m_1}{m}\right)^k \left(\frac{m_2}{m}\right)^{n-k}$$

Exercice 1.2.8 Dans ce dernier exercice, les boules sélectionnées ne sont pas remises dans l'urne.

1. Vérifier que le n -uplet (X_1, \dots, X_n) est une v.a. à valeurs dans l'espace produit $E = \{1, \dots, m\}^n$, et sa loi est donnée par la formule suivante

$$\mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = 1_D(x_1, \dots, x_n) \frac{1}{(m)_n}$$

avec $D = \{(x_1, \dots, x_n) \in E : \forall i \neq j \quad x_i \neq x_j\}$

2. On note respectivement B_1 , et B_2 , l'ensemble des indices des boules noires, et des boules rouges. Montrer que

$$|D \cap B_1^n| = (m_1)_n \quad \text{et} \quad |D \cap B_2^n| = (m_2)_n$$

Plus généralement, vérifier que pour toute collection d'indices $(i_1, \dots, i_n) \in \{1, 2\}^n$, on a

$$|D \cap (B_{i_1} \times \dots \times B_{i_n})| = (m_1)_{n_1} (m_2)_{n_2} \quad \text{avec} \quad n_1 = \sum_{p=1}^n 1_{\{1\}}(i_p) = n - n_2$$

En déduire les formules

$$\mathbb{P}^{(X_1, \dots, X_n)}(D \cap (B_{i_1} \times \dots \times B_{i_n})) = \frac{(m_1)_{n_1} (m_2)_{n_2}}{(m)_n}$$

3. On pose $S_n = \sum_{k=1}^n 1_{B_1}(X_k)$, vérifier que la v.a. S_n est distribuée sur $\{0, \dots, n\}$ selon la loi

$$\mathbb{P}^{S_n}(k) = C_n^k \times \frac{(m_1)_k (m_2)_{n-k}}{(m)_n} = \frac{C_{m_1}^k \times C_{m_2}^{n-k}}{C_m^n}$$

1.2.4 L'indépendance

Définition 1.2.5 Deux v.a. $X_1 : \Omega \rightarrow E_1$, et $X_2 : \Omega \rightarrow E_2$, définies sur un même espace probabilisé (Ω, \mathbb{P}) , sont dites indépendantes, lorsque l'on a pour tout $(x_1, x_2) \in (E_1 \times E_2)$

$$\mathbb{P}^{(X_1, X_2)}(x_1, x_2) = \mathbb{P}^{X_1}(x_1) \times \mathbb{P}^{X_2}(x_2)$$

Plus généralement, n v.a. $X_i : \Omega \rightarrow E_i$, avec $1 \leq i \leq n$, sont dites indépendantes, lorsque l'on a pour tout n -uplet $(x_1, \dots, x_n) \in (E_1 \times \dots \times E_n)$

$$\mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \mathbb{P}^{X_1}(x_1) \times \dots \times \mathbb{P}^{X_n}(x_n)$$

L'égalité précédente revient à dire que la loi du n -uplet (X_1, \dots, X_n) est le produit tensoriel des lois des v.a. X_i ; on utilise parfois la notation synthétique suivante

$$\mathbb{P}^{(X_1, \dots, X_n)} = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}$$

On notera que cette condition est équivalente au fait que pour tout $1 \leq k \leq n$, et toute collection d'indices $1 \leq i_1 < \dots < i_k \leq n$,

$$\mathbb{P}^{(X_{i_1}, \dots, X_{i_k})} = \mathbb{P}^{X_{i_1}} \otimes \dots \otimes \mathbb{P}^{X_{i_k}}$$

Définition 1.2.6 Deux évènements $A_1, A_2 \in \mathcal{F}$ sont dits indépendants lorsque les v.a. indicatrices $(X_1, X_2) = (1_{A_1}, 1_{A_2})$ sont indépendantes.

On notera que

$$\begin{aligned} A_1, A_2 \text{ indépendants} &\Leftrightarrow \mathbb{P}(A_1 \cap A_2) = \mathbb{P}^{(X_1, X_2)}(\{(1, 1)\}) \\ &= \mathbb{P}^{X_1}(\{1\}) \times \mathbb{P}^{X_2}(\{1\}) \\ &= \mathbb{P}^{X_1}(1) \times \mathbb{P}^{X_2}(1) \\ &= \mathbb{P}(A_1) \times \mathbb{P}(A_2) \end{aligned}$$

De même que

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2^c) &= \mathbb{P}^{(X_1, X_2)}(\{(1, 0)\}) = \mathbb{P}^{X_1}(\{1\}) \mathbb{P}^{X_2}(\{0\}) = \mathbb{P}(A_1) \mathbb{P}(A_2^c) \\ \mathbb{P}(A_1^c \cap A_2) &= \mathbb{P}^{(X_1, X_2)}(\{(0, 1)\}) = \mathbb{P}^{X_1}(\{0\}) \mathbb{P}^{X_2}(\{1\}) = \mathbb{P}(A_1^c) \mathbb{P}(A_2) \\ \mathbb{P}(A_1^c \cap A_2^c) &= \mathbb{P}^{(X_1, X_2)}(\{(0, 0)\}) = \mathbb{P}^{X_1}(\{0\}) \mathbb{P}^{X_2}(\{0\}) = \mathbb{P}(A_1^c) \mathbb{P}(A_2^c) \end{aligned}$$

Cependant, on a

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2) &= \mathbb{P}^{(X_1, X_2)}(\{(1, 0), (0, 1), (1, 1)\}) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)\end{aligned}$$

Définition 1.2.7 Les évènements A_1, \dots, A_n , sont dits indépendants, si pour tout $1 \leq k \leq n$, et $1 \leq i_1 < \dots < i_k \leq n$, les v.a.

$$(X_{i_1}, \dots, X_{i_k}) = (1_{A_{i_1}}, \dots, 1_{A_{i_k}})$$

sont indépendantes, ou encore de façon équivalente lorsque l'on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_k})$$

Exemple 1.2.9 Comme son nom l'indique, la notion d'indépendance reflète l'absence de toute interaction entre les réalisations de phénomènes aléatoires.

1. Les v.a. (X_1, \dots, X_n) représentant les résultats successifs de lancers de pièces de monnaie, ou encore celles associées aux montées et descentes lors d'une évolution de particule, étudiées dans les exemples 1.2.6, et 1.2.7, forment des collections de v.a. indépendantes.
2. Les v.a. (X_1, \dots, X_n) représentant les numéros des boules sélectionnées avec remise dans une urne, et étudiées dans l'exercice 1.2.6, sont indépendantes.
3. Les indicatrices $X_i = 1_{A_i}$ d'évènements indépendants A_i , forment des séquences de v.a. indépendantes.

La notion d'indépendance entre v.a. permet d'élaborer une multitude de modèles probabilistes, sans passer par une description précise de l'espace de probabilité sur lequel les v.a. se réalisent, mais plutôt fondés sur de nouvelles interprétations probabilistes. A titre d'exemple, nous verrons très bientôt que la loi binomiale peut s'interpréter comme la somme v.a. élémentaires de Bernoulli.

Cette notion d'indépendance est donc très naturelle et intuitive. Néanmoins, dans certaines situations, certains jugements hâtifs peuvent s'avérer catastrophiques! Nous laissons le soin au lecteur d'apprécier cette observation, en s'exerçant sur les trois exercices suivants.

Exercice 1.2.9 On considère un espace d'évènements à quatre états

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

muni de la probabilité uniforme $\mathbb{P}(\omega_i) = 1/4$. On considère les évènements suivants

$$A_1 = \{\omega_1, \omega_2\} \quad A_2 = \{\omega_2, \omega_3\} \quad A_3 = \{\omega_1, \omega_4\}$$

Vérifier que ces trois évènements sont deux à deux indépendants, mais l'on a néanmoins

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) \neq \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \mathbb{P}(A_3)$$

Exercice 1.2.10 Soit $\Omega = \{1, \dots, 6\}^2$, l'espace des épreuves correspondant au lancer de deux dés. On considère les évènements suivants.

$$\begin{aligned} A_1 &= \{(i, j) \in \Omega : j \in \{1, 2, 5\}\} & A_2 &= \{(i, j) \in \Omega : j \in \{4, 5, 6\}\} \\ A_3 &= \{(i, j) \in \Omega : i + j = 9\} = \{(3, 6), (6, 3), (4, 5), (5, 4)\} \end{aligned}$$

1. Vérifier que

$$\begin{aligned} A_1 \cap A_2 &= \{(i, j) \in \Omega : j = 5\} & A_1 \cap A_3 &= \{(4, 5)\} \\ A_2 \cap A_3 &= \{(4, 5), (3, 6), (5, 4)\} & A_1 \cap A_2 \cap A_3 &= \{(4, 5)\} \end{aligned}$$

2. Montrer que

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \mathbb{P}(A_3)$$

mais, pour tout $i \neq j$

$$\mathbb{P}(A_i \cap A_j) \neq \mathbb{P}(A_i) \times \mathbb{P}(A_j)$$

Exercice 1.2.11 On considère un modèle d'urne contenant m_1 boules de couleur c_1 , m_2 boules de c_2, \dots , et m_r boules de couleur c_r . On sélectionne au hasard n boules avec remise, et l'on note X_1, \dots, X_n leurs couleurs.

1. Vérifier que le n -uplet (X_1, \dots, X_n) peut être réalisé sur l'espace des suites

$$\omega = (\omega_1, \dots, \omega_n) \in \Omega = \{c_1, \dots, c_r\}^n$$

muni de la probabilité uniforme. Vérifier que l'on a pour tout n -uplet d'indices $(i_1, \dots, i_n) \in \{1, \dots, r\}^n$

$$\mathbb{P}^{(X_1, \dots, X_n)}(c_{i_1}, \dots, c_{i_n}) = \prod_{p=1}^n \frac{m_{i_p}}{m} \quad \text{avec} \quad m = \sum_{i=1}^r m_i$$

2. Vérifier que l'on a pour tout $1 \leq p \leq n$, et pour tout $i \in \{1, \dots, r\}$

$$\mathbb{P}^{X_p}(c_i) = m_i/m$$

En conclure que les v.a. X_i sont indépendantes.

Exercice 1.2.12 On considère une suite d'évènements indépendants A_1, \dots, A_n , se déroulant au cours du temps, avec la même probabilité de réalisation

$$p = \mathbb{P}(A_i)$$

On associe à cette suite, la v.a. $X = \sum_{i=1}^n 1_{A_i}$ représentant le nombre de fois où les évènements se réalisent.

1. Montrer que pour tout $I \subset \{1, \dots, n\}$, tel que $|I| = k$, on a

$$\mathbb{P}([\cap_{i \in I} A_i] \cap [\cap_{i \notin I} A_i^c]) = p^k (1-p)^{n-k}$$

2. En déduire que X est distribuée sur $E = \{0, \dots, n\}$, selon la loi

$$\mathbb{P}^X(k) = C_n^k p^k (1-p)^{n-k}$$

Exercice 1.2.13 Soient X_1, \dots, X_n une suite de n copies indépendantes d'une v.a. de Bernoulli X sur $\{0, 1\}$

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p \in [0, 1]$$

1. Vérifier que pour tout n -uplet $(x_1, \dots, x_n) \in \{0, 1\}^n$, on a

$$\mathbb{P}^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}$$

2. Quelle est la loi de la v.a. de comptage $S_n = \sum_{i=1}^n X_i$.

1.3 Le conditionnement

1.3.1 Évènements

L'information la plus élémentaire que l'on puisse considérer correspond à la donnée d'un évènement simple ; tel la connaissance de la parité d'un lancer de dés, une région précise visitée par une évolution aléatoire, la valeur d'un actif boursier à un date donnée, ou encore le niveau de saturation d'un réseau de communication. Pour prédire d'autres évènements connexes à une telle information, il est nécessaire de restreindre toutes les situations envisageables à cette donnée. Cette restriction événementielle se traduit mathématiquement par une simple intersection ensembliste. Cette discussion nous amène à introduire les notions suivantes.

Définition 1.3.1 Soit (Ω, \mathbb{P}) un espace de probabilités. La probabilité d'un évènement $A \subset \Omega$, conditionnelle en un évènement $B \subset \Omega$, tel que $\mathbb{P}(B) > 0$, est la quantité définie par

$$\mathbb{P}(A | B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$$

Lorsque $\mathbb{P}(B) = 0$, on pose $\mathbb{P}(A | B) = 0$. Si X désigne une v.a. réelle (à valeurs dans un espace fini E), on définit l'espérance conditionnelle de X en l'évènement B , avec $\mathbb{P}(B) > 0$, par la quantité

$$\mathbb{E}(X | B) = \mathbb{E}(X \mathbf{1}_B) / \mathbb{E}(\mathbf{1}_B) = \mathbb{E}(X \mathbf{1}_B) / \mathbb{P}(B)$$

Lorsque $\mathbb{P}(B) = 0$, on pose $\mathbb{E}(X | B) = 0$.

Pour des variables indicatrices d'évènements $X = \mathbf{1}_A$, avec $A \subset \Omega$, ces deux notions coïncident $\mathbb{E}(\mathbf{1}_A | B) = \mathbb{P}(A | B)$.

L'espérance conditionnelle en un évènement hérite des propriétés de linéarité des espérances simples. Plus précisément, pour tout couple de v.a. sont réelles (X_1, X_2) , nous avons la propriété de linéarité

$$\forall (a_1, a_2) \in \mathbb{R}^2 \quad \mathbb{E}(a_1 X_1 + a_2 X_2 \mid B) = a_1 \mathbb{E}(X_1 \mid B) + a_2 \mathbb{E}(X_2 \mid B)$$

Avant de développer une série de propriétés, commençons par visualiser ces deux objets sur un exemple élémentaire.

Exemple 1.3.1 Dans l'exemple du jet de dés uniforme sur $\Omega = \{1, 2, \dots, 6\}$, on a

$$\mathbb{P}(\{2\} \mid \{2, 4, 6\}) = \frac{1/6}{3/6} = 1/3 \quad \text{et} \quad \mathbb{P}(\{2\} \mid \{1, 3, 5\}) = 0$$

On notera que le résultat du lancer de dés peut se représenter par la v.a. identité

$$X : \omega \in \Omega = \{1, \dots, 6\} \mapsto X(\omega) = \omega \in \mathbb{R}$$

Dans ce cas, nous avons

$$\begin{aligned} \mathbb{E}(X \mid \{2, 4, 6\}) &= \sum_{x \in \{2, 4, 6\}} x \mathbb{P}(\{x\} \mid \{2, 4, 6\}) = (2 + 4 + 6)/3 = 4 \\ \mathbb{E}(X \mid \{1, 3, 5\}) &= \sum_{x \in \{1, 3, 5\}} x \mathbb{P}(\{x\} \mid \{1, 3, 5\}) = (1 + 3 + 5)/3 = 3 \end{aligned}$$

Il est important de souligner que ces notions de conditionnement sont en adéquation parfaite avec l'intuition et l'expérience. Sachant que le résultat du lancer de dés est l'une des trois valeurs 2, 4, ou 6, il y a bien évidemment une chance sur 3 que le résultat soit le chiffre 2. La moyenne du lancer sera dans ce cas, le barycentre 4 des trois valeurs possibles $\{2, 4, 6\}$.

1. [**Évènements ordonnés.**] Lorsque $A \subset B$, on dit que la réalisation de l'évènement A entraîne celle de B , ou plus formellement " $A \Rightarrow B$ ", en ce sens où $\omega \in A \Rightarrow \omega \in B$, soit encore en terme d'indicatrices $1_A \leq 1_B$. Cette interprétation est très claire dans l'exemple du lancer de dés. L'inclusion $\{2, 4\} \subset \{2, 4, 6\}$ témoigne du fait que le résultat du lancer sera nécessairement pair si les chiffres 2 ou 4 se réalisent. De même que la réalisation des chiffres 2 ou 4 est bien plus probable si l'on sait par avance que le résultat du lancer sera pair. Le résultat qui suit est aussi intuitivement très clair

$$A \subset B \implies \mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B) \geq \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(B|A) = 1$$

La première assertion montre que la probabilité a posteriori $\mathbb{P}(A|B)$ d'un évènement A entraînant un évènement B est toujours plus grande que sa probabilité de réalisation a priori $\mathbb{P}(A)$. Autrement dit, la connaissance d'un évènement B nécessaire à la réalisation de A apporte toujours de l'information. Autrement dit, la donnée de l'évènement B rend plus probable la réalisation de A . La seconde assertion s'interprète de la façon suivante. Si la réalisation de A entraîne celle de B ; lorsque A se réalise, il en est de même de B .

2. [**Indépendance et conditionnement.**] Lorsque les évènements A , et B , sont indépendants, on note que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) \implies \mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B)$$

3. [Conditionnement uniforme.]

Pour des probabilités uniformes \mathbb{P} sur des espaces mesurables discrets (Ω, \mathcal{F}) , on a

$$\mathbb{P}(A) = |A|/|\Omega| \Rightarrow \mathbb{P}(A | B) = |A \cap B|/|B|$$

Dans cette situation, tout se passe comme si l'on "restreint" toute l'expérience à l'évènement B . Dans ce contexte, la notion de conditionnement est donc clairement liée à l'acquisition d'une nouvelle information sur l'expérience : "l'évènement B se réalise".

4. [Formule de Bayes.]

Soit $(A_i)_{i \in I}$ une partition (finie) de Ω , et soit B un évènement de probabilité non nulle $\mathbb{P}(B) > 0$. La formule de Bayes permet de transformer une suite de probabilités a priori $(\mathbb{P}(A_i))_{i \in I}$, en une suite de probabilités a posteriori $(\mathbb{P}(A_i|B))_{i \in I}$

$$\forall i \in I \quad \mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}(B|A_j) \mathbb{P}(A_j)}$$

Preuve:

On vérifie aisément ce résultat en utilisant la symétrie de la formule

$$\begin{aligned} \mathbb{P}(A_i \cap B) &= \mathbb{P}(A_i|B)\mathbb{P}(B) = \mathbb{P}(B|A_i)\mathbb{P}(A_i) \\ &\Downarrow \\ \mathbb{P}(A_i|B) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}(B|A_i) \mathbb{P}(A_i) \end{aligned}$$

La forme additive de la constante de normalisation, résulte simplement du fait que

$$\sum_{j \in I} \mathbb{P}(A_j|B) = 1 \Rightarrow \mathbb{P}(B) = \sum_{j \in I} \mathbb{P}(B|A_j) \mathbb{P}(A_j)$$

■

L'exercice suivant offre une initiation à l'analyse statistique d'un problème pratique. Le lecteur est vivement encouragé à décrire un espace de probabilité précis rendant compte des événements en question. Pour développer son intuition probabiliste, on pourra s'exercer à trouver (ou retrouver) les réponses sans faire de calcul!

Exercice 1.3.1 Deux machines industrielles \mathcal{M}_1 et \mathcal{M}_2 ont des taux de production quotidienne d'objets défectueux égaux à $p_1 = 5\%$ et $p_2 = 10\%$. Chacune de ces machines \mathcal{M}_1 et \mathcal{M}_2 produit respectivement $m_1 = 100$, et $m_2 = 200$ objets. Qu'elle est la probabilité pour qu'un objet pris au hasard soit défectueux? Qu'elle est la probabilité pour que ce soit la première machine \mathcal{M}_1 qui l'ait produit?

L'exercice suivant possède une structure temporelle intéressante, permettant notamment de guider l'intuition probabiliste, et trouver ainsi une solution élémentaire. Cette étude est par essence assez proche de la notion de processus de Markov que nous aborderons à la fin de ce cours.

Exercice 1.3.2 *Un sac contient 2 pièces de monnaie. L'une équilibrée, et ayant une probabilité 1/2 de donner "pile" ou "face"; la seconde ayant une probabilité 1/3 de donner "face". On lance au hasard l'une des pièces, et l'on observe un résultat "face". Quelle est la probabilité d'avoir choisi la pièce équilibrée ?*

Exercice 1.3.3 (Formule de Bayes séquentielle) *Montrer que la probabilité conditionnelle d'une réalisation conjointe de n évènements $(A_p)_{1 \leq p \leq n}$ par rapport à un évènement B (de probabilité non nulle) est donnée par la formule multiplicative*

$$\begin{aligned} & \mathbb{P}(\cap_{p=1}^n A_p | B) \\ &= \mathbb{P}(A_n | B \cap [\cap_{p=1}^{n-1} A_p]) \mathbb{P}(A_{n-1} | B \cap [\cap_{p=1}^{n-2} A_p]) \dots \mathbb{P}(A_2 | B \cap A_1) \mathbb{P}(A_1 | B) \\ &= \prod_{p=1}^n \mathbb{P}(A_p | B \cap [\cap_{q=1}^{p-1} A_q]) \end{aligned}$$

avec la convention $\prod_{\emptyset} = \Omega$, lorsque $p = 1$.

Exercice 1.3.4 *Soient X_1, \dots, X_n une suite de n copies indépendantes d'une v.a. de Bernoulli X sur $\{0, 1\}$*

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p \in [0, 1]$$

On note $S_k = \sum_{i=1}^k X_i$, les variables de comptage du nombre de succès. Montrer que pour tout $m \leq n$, et $0 \leq k \leq l \leq n$, on a

$$\mathbb{P}^{S_m | S_n}(k | l) = \frac{C_{n-m}^{l-k} C_m^k}{C_n^l} \quad \text{et} \quad \mathbb{P}^{X_1 | S_n}(1 | l) = \frac{l}{n}$$

1.3.2 Décompositions et sous algèbres

La connaissance d'une famille d'évènements disjoints apporte bien entendu une information plus fine que celle liée à un seul ! C'est le cas de la parité ou non d'un lancer de dés, l'échec ou le succès d'une expérience, ou encore la donnée du niveau de saturation d'une file d'attente. Pour traiter l'information contenue dans chaque évènement, on est amené à introduire un nouveau cadre fonctionnel pour détecter quel évènement "s'est réalisé".

Définition 1.3.2 Soit $\mathcal{D} = (D_i)_{i \in I}$ une partition dénombrable d'un espace fini Ω . Autrement dit, \mathcal{D} est une décomposition de l'espace Ω en une suite d'évènements disjoints, c'est à dire

$$\forall i > j \geq 1 \quad D_i \cap D_j = \emptyset, \quad \text{et} \quad \Omega = \cup_{i \in I} D_i$$

La probabilité conditionnelle $\mathbb{P}(\cdot | \mathcal{D})$ par rapport à \mathcal{D} est la fonction (mesurable) sur $(\Omega \times \mathcal{P}(\Omega))$ définie par

$$(\omega, A) \in (\Omega \times \mathcal{P}(\Omega)) \mapsto \mathbb{P}(A | \mathcal{D})(\omega) = \sum_{i \in I} \mathbb{P}(A | D_i) 1_{D_i}(\omega)$$

L'espérance conditionnelle d'une v.a. réelle X par rapport à une décomposition \mathcal{D} est la fonction donnée par par

$$\mathbb{E}(X | \mathcal{D})(\omega) = \sum_{i \in I} \mathbb{E}(X | D_i) 1_{D_i}(\omega)$$

Pour des variables indicatrices $X = 1_A$, $A \subset \Omega$, on a $\mathbb{E}(1_A | \mathcal{D}) = \mathbb{P}(A | \mathcal{D})$. Enfin, on notera que les fonctions

$$\mathbb{E}(X | \mathcal{D}) : \omega \in \Omega \mapsto \mathbb{E}(X | \mathcal{D})(\omega)$$

sont mesurables non seulement par rapport à l'algèbre discrète de référence $\mathcal{P}(\Omega)$, mais aussi par rapport à l'algèbre $\mathcal{F} = \sigma(\mathcal{D})$ ($\subset \mathcal{P}(\Omega)$) engendrée par la décomposition \mathcal{D} . Pour vérifier cette assertion, il suffit de noter que

$$\begin{aligned} \mathbb{E}(X | \mathcal{D})^{-1}(A) &= \{\omega \in \Omega : \mathbb{E}(X | \mathcal{D})(\omega) \in A\} \\ &= \{\omega \in \Omega : \sum_{i \in I} \mathbb{E}(X | D_i) 1_{D_i}(\omega) \in A\} = \cup_{i \in I : \mathbb{E}(X | D_i) \in A} D_i \in \mathcal{F} = \sigma(\mathcal{D}) \end{aligned}$$

Pour souligner cette propriété on utilise souvent les notations suivantes :

$$\mathcal{F} = \sigma(\mathcal{D}) \implies \mathbb{E}(X | \mathcal{F}) = \mathbb{E}(X | \mathcal{D})$$

Avec ces notations, et dans le cas de l'algèbre triviale $\{\emptyset, \Omega\}$ (associée à la partition $\mathcal{D} = \{\Omega\}$), on notera que

$$\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X|\Omega) = \mathbb{E}(X)$$

Il est à nouveau important de souligner que les objets fonctionnels que nous venons de définir héritent des propriétés d'additivité, et de linéarité des probabilités et des espérances conditionnelles par rapport à un événement. Plus précisément, pour tout couple de v.a. sont réelles (X_1, X_2) , nous avons la propriété de linéarité

$$\forall (a_1, a_2) \in \mathbb{R}^2 \quad \mathbb{E}(a_1 X_1 + a_2 X_2 | \mathcal{D}) = a_1 \mathbb{E}(X_1 | \mathcal{D}) + a_2 \mathbb{E}(X_2 | \mathcal{D})$$

Exemple 1.3.2 Reprenons l'exemple du jet de dés uniforme décrit dans l'exemple 1.3.1, avec la partition $\mathcal{D} = \{\{2, 4, 6\}, \{1, 3, 5\}\}$ de $\Omega = \{1, \dots, 6\}$ en événements pairs ou impairs. Dans cette situation, nous avons

$$\begin{aligned} \mathbb{P}(\{2\} | \mathcal{D}) &= \mathbb{P}(\{2\} | \{2, 4, 6\}) 1_{\{2,4,6\}} + \mathbb{P}(\{2\} | \{1, 3, 5\}) 1_{\{1,3,5\}} \\ &= \frac{1}{3} 1_{\{2,4,6\}} \end{aligned}$$

et $\mathbb{P}(\{1, 3\} | \mathcal{D}) = \frac{2}{3} 1_{\{1,3,5\}}$. Ces décompositions fonctionnelles s'interprètent de la façon suivante. Si le résultat de l'expérience est pair, autrement dit si l'évènement $\{2, 4, 6\}$ se réalise, alors la probabilité d'avoir le chiffre 2 est de $1/3$

$$\omega \in \{2, 4, 6\} \implies \mathbb{P}(\{2\} | \mathcal{D})(\omega) = 1/3$$

Inversement, si le résultat de l'expérience est impair, autrement dit si l'évènement $\{1, 3, 5\}$ se réalise, alors la probabilité d'avoir le chiffre 2 est nulle. Autrement dit, nous avons

$$\omega \in \{1, 3, 5\} \implies \mathbb{P}(\{2\} | \mathcal{D})(\omega) = 0$$

Avec les notations de l'exemple 1.3.1, on obtient

$$\begin{aligned} \mathbb{E}(X | \mathcal{D}) &= \mathbb{E}(X | \{2, 4, 6\}) 1_{\{2,4,6\}} + \mathbb{E}(X | \{1, 3, 5\}) 1_{\{1,3,5\}} \\ &= 4 1_{\{2,4,6\}} + 3 1_{\{1,3,5\}} \end{aligned}$$

Dans cette exemple, la propriété de $\sigma(\mathcal{D})$ -mesurabilité soulignée précédemment est assez simple à voir. On notera que

$$\{\omega : \mathbb{P}(\{2\}|\mathcal{D}) = 1/3\} = \{2, 4, 6\} = \{\omega : \mathbb{E}(X|\mathcal{D}) = 4\}$$

1. **[Représentations fonctionnelles]**. Lorsque $\mathcal{D} = \{B, B^c\}$ est la décomposition associée à un évènement $B \subset \Omega$, avec $\mathbb{P}(B) \in (0, 1)$, on remarque que

$$\mathbb{P}(A | \mathcal{D})(\omega) = \mathbb{P}(A|B) 1_B(\omega) + \mathbb{P}(A|B^c) 1_{B^c}(\omega)$$

Dans cette situation, l'espérance conditionnelle d'une variable X par rapport à $\mathcal{D} = \{B, B^c\}$ est donnée par la v.a.

$$\mathbb{E}(X | \mathcal{D})(\omega) = \mathbb{E}(X|B) 1_B(\omega) + \mathbb{E}(X|B^c) 1_{B^c}(\omega)$$

Pour la partition triviale $\mathcal{D} = \{\emptyset, \Omega\}$, on a bien entendu la formule suivante

$$\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X|\Omega) = \mathbb{E}(X)$$

2. [Décompositions ordonnées].

Définition 1.3.3 Soit $\mathcal{D}' = (D'_i)_{i \in I'}$ et $\mathcal{D} = (D_i)_{i \in I}$ un couple de partitions sur Ω . On dit que \mathcal{D}' est plus fine que \mathcal{D} , ou encore \mathcal{D} est plus grossière que \mathcal{D}' , et on note $\mathcal{D} \subset \mathcal{D}'$, lorsque tout élément de \mathcal{D} s'exprime comme la réunion d'éléments de \mathcal{D}'

$$\mathcal{D} \subset \mathcal{D}' \iff (\forall i \in I \quad \exists I_i \subset I' \quad D_i = \cup_{j \in I_i} D'_j)$$

Cette propriété induit relation d'ordre partiel sur l'ensemble des partitions.

Dans cette situation, et d'après la formule de factorisation, on a

$$D_i \cap D'_j = (\cup_{k \in I_i} D'_k) \cap D'_j = \cup_{k \in I_i} (D'_k \cap D'_j) = 1_{I_i}(j) D'_j$$

Par conséquent, on obtient $\mathbb{P}(D_i|D'_j) = 1_{I_i}(j)$.

On montre alors aisément la formule de conditionnements emboîtés

$$\mathcal{D} \subset \mathcal{D}' \implies \mathbb{E}(\mathbb{E}(X | \mathcal{D}') | \mathcal{D}) = \mathbb{E}(X | \mathcal{D})$$

valable pour toute v.a. réelle. En particulier, pour $\mathcal{D} = \{\emptyset, \Omega\}$, on trouve que

$$\mathbb{E}(\mathbb{E}(X | \mathcal{D}')) = \mathbb{E}(X)$$

Preuve:

On commence par noter que

$$\begin{aligned} \mathbb{E}(\mathbb{E}(X | \mathcal{D}') | D_i) &= \mathbb{E}\left(\sum_{j \in I'} \mathbb{E}(X | D'_j) 1_{D'_j} | D_i\right) \\ &= \sum_{j \in I'} \mathbb{E}(X | D'_j) \mathbb{P}(D'_j \cap D_i) / \mathbb{P}(D_i) \end{aligned}$$

D'après la discussion précédente, on obtient

$$\begin{aligned} \mathbb{E}(\mathbb{E}(X | \mathcal{D}') | D_i) &= \sum_{j \in I'} \mathbb{E}(X 1_{D'_j}) \frac{\mathbb{P}(D'_j \cap D_i)}{\mathbb{P}(D'_j)\mathbb{P}(D_i)} \\ &= \sum_{j \in I'} \mathbb{E}(X 1_{D'_j}) \mathbb{P}(D_i|D'_j) / \mathbb{P}(D_i) \\ &= \sum_{j \in I_i} \mathbb{E}(X 1_{D'_j}) / \mathbb{P}(D_i) = \mathbb{E}(X 1_{D_i}) / \mathbb{P}(D_i) \end{aligned}$$

Par conséquent, on a

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{D}') \mid D_i) = \mathbb{E}(X \mid D_i)$$

On en conclut que

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{D}') \mid \mathcal{D}) = \sum_{i \in I} \mathbb{E}(X \mid D_i) 1_{D_i} = \mathbb{E}(X \mid \mathcal{D})$$

■

3. [**Sous algèbres ordonnées**]. Commençons par montrer qu'une sous algèbre d'évènements $\mathcal{F} \subset \mathcal{P}(\Omega)$ est toujours engendrée par une partition \mathcal{D} de Ω . Autrement dit, pour toute algèbre \mathcal{F} , il existe une unique partition \mathcal{D} de Ω telle que

$$\mathcal{F} = \sigma(\mathcal{D}) = \sigma(\{1_D : D \in \mathcal{D}\})$$

On construit cette partition, en choisissant le plus grand sous-ensemble $\mathcal{D} = (D_i)_{i \in I}$ formé d'évènements deux à deux disjoints de \mathcal{F} (voir exercice 1.2.5).

Soit \mathcal{F}' une sous algèbre de $\mathcal{P}(\Omega)$, plus fine qu'une algèbre donnée \mathcal{F} , en ce sens où

$$\mathcal{F} = \sigma(\mathcal{D}) \subset \mathcal{F}' = \sigma(\mathcal{D}')$$

pour un couple $(\mathcal{D}, \mathcal{D}')$ de partitions de Ω . La formule de conditionnements emboîtés décrite précédemment s'exprime en terme de sous algèbres sous la forme suivante

$$\mathcal{F} \subset \mathcal{F}' \implies \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}') \mid \mathcal{F}) = \mathbb{E}(X \mid \mathcal{F})$$

4. [**Mesurabilité**]. Pour toute v.a. réelle X , et pour toute algèbre $\mathcal{F} = \sigma(\mathcal{D})$, engendrée par une partition \mathcal{D} , nous avons

$$X \text{ } \mathcal{F} \text{-mesurable} \implies \mathbb{E}(X \mid \mathcal{F}) = X$$

Preuve:

On remarque que toute v.a. X , mesurable par rapport à une algèbre $\sigma(\mathcal{D})$, peut se mettre sous la forme discrète $X = \sum_{i \in I} x_i 1_{D_i}$, pour certains nombres réels x_i , tels que $D_i = X^{-1}(\{x_i\})$. Avec ces notations, on obtient

$$\begin{aligned} \mathbb{E}(X \mid \mathcal{D}) &= \sum_{i \in I} x_i \mathbb{P}(D_i \mid \mathcal{D}) = \sum_{i \in I} x_i \left[\sum_{j \in I} \mathbb{P}(D_i \mid D_j) 1_{D_j} \right] \\ &= \sum_{i \in I} x_i \left[\sum_{j \in I} 1_i(j) 1_{D_j} \right] = \sum_{i \in I} x_i 1_{D_i} = X \end{aligned}$$

■

1.3.3 Variables aléatoires

La notion de conditionnement entre v.a. discrètes est par essence “assez proche” de la notion de conditionnement par rapport à une décomposition de l’espace des évènements. Pour s’en convaincre, on notera que les évènements associés aux différentes valeurs prises par une v.a. discrète forment une partition de l’espace des évènements élémentaires, *et inversement*. Afin de souligner les liens entre cette nouvelle notion et les processus aléatoires, nous avons choisi d’indicer les v.a. par des nombres entiers. Dans la suite du cours, lorsque nous aborderons la notion de processus, ces indices entiers correspondront aux paramètres temporels.

Définition 1.3.4 *On se donne un couple de v.a. (X_1, X_2) à valeurs dans un espace produit fini $(E_1 \times E_2)$, muni de la tribu discrète $\mathcal{P}(E_1 \times E_2)$. On supposera que ces variables sont définies sur un espace probabilisé (Ω, \mathbb{P}) . La loi conditionnelle de X_2 en X_1 , est la collection de lois de probabilités $\mathbb{P}^{X_2|X_1}(\cdot | x_1)$ sur $(E_2, \mathcal{P}(E_2))$, indexée par les $x_1 \in E_1$ et données par*

$$x_2 \in E_2 \mapsto \mathbb{P}^{X_2|X_1}(x_2|x_1) = \mathbb{P}(X_2 = x_2 | X_1 = x_1) \in [0, 1]$$

pour tout x_1 tel que $\mathbb{P}^{X_1}(x_1) > 0$. Lorsque $\mathbb{P}^{X_1}(x_1) = 0$, on convient que $\mathbb{P}^{X_2|X_1}(x_2|x_1) = 0$.

Lorsque $\mathbb{P}^{X_1}(x_1) = 0$, on observe que

$$\mathbb{P}^{X_1}(x_1) = \sum_{x_2 \in E_2} \mathbb{P}^{X_1, X_2}(x_1, x_2) = 0 \Rightarrow \forall x_2 \in E_2 \quad \mathbb{P}^{X_1, X_2}(x_1, x_2) = 0$$

Notre convention ($\mathbb{P}^{X_1}(x_1) = 0 \Rightarrow \mathbb{P}^{X_2|X_1}(x_2|x_1) = 0$), est de ce fait naturelle, et cohérente. Lorsque les v.a. sont égales $X_1 = X_2$, on a clairement

$$\mathbb{P}^{X_1|X_1}(x_2|x_1) = \frac{\mathbb{P}(X_1 = x_1, X_1 = x_2)}{\mathbb{P}(X_1 = x_1)} = 1_{x_1}(x_2)$$

pour tout $\mathbb{P}^{X_1}(x_1) > 0$. On notera enfin la formule de conditionnement

$$\mathbb{P}^{(X_1, X_2)}(x_1, x_2) = \mathbb{P}^{X_2|X_1}(x_2|x_1) \mathbb{P}^{X_1}(x_1) \quad (1.3)$$

Exemple 1.3.3 *Le jeu du lancer de dés étudié dans l'exemple 1.3.1 peut être décrit par le couple de variables (X_1, X_2) sur $\Omega = \{1, \dots, 6\}$ définies par*

$$X_1(\omega) = 1_{\{2,4,6\}}(\omega) \quad \text{et} \quad X_2(\omega) = \omega$$

La v.a. X_2 représente le résultat de l'expérience, et la v.a. X_1 la parité du résultat. On notera que $\{X_2 = i\} = \{i\}$, pour chaque $i = 1, \dots, 6$, alors que

$$\{X_1 = 1\} = \{2, 4, 6\} \quad \text{et} \quad \{X_1 = 0\} = \{1, 3, 5\}$$

On obtient ainsi immédiatement les formules

$$\begin{aligned} \forall i \in \{2, 4, 6\} \quad \mathbb{P}^{X_2|X_1}(i|1) &= 1/3 \quad \text{et} \quad \mathbb{P}^{X_2|X_1}(i|0) = 0 \\ \forall i \in \{1, 3, 5\} \quad \mathbb{P}^{X_2|X_1}(i|1) &= 0 \quad \text{et} \quad \mathbb{P}^{X_2|X_1}(i|0) = 1/3 \end{aligned}$$

On remarque que pour toute fonction f sur $(E_1 \times E_2)$

$$\begin{aligned} \mathbb{E}(f(X_1, X_2)) &= \sum_{(x_1, x_2) \in (E_1 \times E_2)} f(x_1, x_2) \mathbb{P}^{X_2|X_1}(x_2|x_1) \mathbb{P}^{X_1}(x_1) \\ &= \sum_{x_1 \in E_1} \left[\sum_{x_2 \in E_2} f(x_1, x_2) \mathbb{P}^{X_2|X_1}(x_2|x_1) \right] \mathbb{P}^{X_1}(x_1) \end{aligned} \quad (1.4)$$

Définition 1.3.5 La quantité entre crochet dans la formule (1.4) correspond à l'espérance de la v.a. $f(X_1, X_2)$ en l'évènement $\{X_1 = x_1\}$

$$\mathbb{E}(f(X_1, X_2)|X_1 = x_1) = \sum_{x_2 \in E_2} f(x_1, x_2) p^{X_2|X_1}(x_2|x_1) dx_2$$

En particulier, lorsque X_2 est une v.a. réelle on a

$$\mathbb{E}(X_2 | X_1 = x_1) = \sum_{x_2 \in E_2} x_2 \mathbb{P}^{X_2|X_1}(x_2|x_1)$$

Pour des fonctions indicatrices $f = 1_{A_2}$, avec $A_2 \subset E_2$, on obtient

$$\mathbb{E}(1_{A_2}(X_2) | X_1 = x_1) = \mathbb{P}^{X_2|X_1}(A_2|x_1)$$

On vérifie sans trop de peine que l'ensemble des évènements

$$D_{x_1} = X_1^{-1}(\{x_1\}) = \{\omega \in \Omega : X_1(\omega) = x_1\} \subset \Omega$$

lorsque x_1 parcourt E_1 , est une partition de Ω . Cette partition est dite engendrée par la variable X_1 , et on la note $\mathcal{D}^{X_1} = (D_{x_1})_{x_1 \in E_1}$. On notera que toute partition $\mathcal{D} = (D_i)_{i \in I}$ de Ω est engendrée par une variable aléatoire X . Il suffit de poser

$$X = \sum_{i \in I} i 1_{D_i}$$

Dans cette situation, la v.a. X est à valeurs dans I , et on a $D_i = X^{-1}(\{i\})$, pour chaque $i \in I$; autrement dit, nous avons $\mathcal{D} = \mathcal{D}^X$.

Avec ces notations, on retrouve les formules de conditionnement par rapport à des décompositions, introduites dans la section précédente :

$$\begin{aligned} \mathbb{P}(X_2 = x_2 | \mathcal{D}^{X_1})(\omega) &= \sum_{x_1 \in E_1} \mathbb{P}^{X_2|X_1}(x_2|x_1) 1_{X_1^{-1}(\{x_1\})}(\omega) \\ \mathbb{E}(f(X_2) | \mathcal{D}^{X_1})(\omega) &= \sum_{x_1 \in E_1} \mathbb{E}(f(X_2)|X_1 = x_1) 1_{X_1^{-1}(\{x_1\})}(\omega) \end{aligned}$$

Comme précédemment, on notera que les fonctions

$$\mathbb{E}(f(X_2) | \mathcal{D}^{X_1}) : \omega \in \Omega \mapsto \mathbb{E}(f(X_2) | \mathcal{D}^{X_1})(\omega)$$

sont mesurables non seulement par rapport à l'algèbre discrète $\mathcal{P}(\Omega)$, mais aussi par rapport à la sous algèbre $\sigma(X_1) = \sigma(\mathcal{D}^{X_1}) \subset \mathcal{F}$, engendrée par la décomposition \mathcal{D}^{X_1} ; ou de manière équivalente, par la variable X_1 . En pratique, les deux formules précédentes sont d'un usage très courant, et elles sont souvent notées

$$\mathbb{E}(f(X_2) | X_1) = \mathbb{E}(f(X_2) | \sigma(X_1)) = \mathbb{E}(f(X_2) | \mathcal{D}^{X_1})$$

Il est à nouveau important de souligner que *les objets fonctionnels que nous venons de définir héritent des propriétés de linéarité des probabilités et des espérances conditionnelles par rapport à une décomposition.*

Exemple 1.3.4 On reprenant l'énoncé et les notations de l'exemple 1.3.3, nous avons $\sigma(X_1) = \{\Omega, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$. Pour toute fonction f sur \mathbb{R} , on obtient la formule

$$\begin{aligned} & \mathbb{E}(f(X_2) | \sigma(X_1)) \\ &= \mathbb{E}(f(X_2) | X_1 = 1) 1_{\{2,4,6\}} + \mathbb{E}(f(X_2) | X_1 = 0) 1_{\{1,3,5\}} \\ &= [f(2) + f(4) + f(6)] \frac{1}{3} 1_{\{2,4,6\}} + [f(1) + f(3) + f(5)] \frac{1}{3} 1_{\{1,3,5\}} \end{aligned}$$

1. **[Formule de Bayes.]** Comme dans le cas des probabilités conditionnelles entre évènements, en utilisant la symétrie de la formule (1.3), on a

$$\mathbb{P}^{X_2 | X_1}(x_2 | x_1) \mathbb{P}^{X_1}(x_1) = \mathbb{P}^{X_1 | X_2}(x_1 | x_2) \mathbb{P}^{X_2}(x_2)$$

D'autre part, on remarque que

$$\mathbb{P}^{X_1}(x_1) = \sum_{(x_1, x_2) \in (E_1 \times E_2)} \mathbb{P}^{X_1, X_2}(x_1, x_2) = \sum_{x_2 \in E_2} \mathbb{P}^{X_1 | X_2}(x_1 | x_2) \mathbb{P}^{X_2}(x_2)$$

Ceci nous conduit assez rapidement à la formule de Bayes

$$\mathbb{P}^{X_2 | X_1}(x_2 | x_1) = \frac{\mathbb{P}^{X_1 | X_2}(x_1 | x_2) \mathbb{P}^{X_2}(x_2)}{\sum_{y_2 \in E_2} \mathbb{P}^{X_1 | X_2}(x_1 | y_2) \mathbb{P}^{X_2}(y_2)}$$

2. **[Formule de Bayes séquentielle.]** La formule de conditionnement (1.3) s'étend à des séquences de v.a. $(X_p)_{0 \leq p \leq n}$ définies sur un même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans des espaces finis $(E_p, \mathcal{P}(E_p))_{0 \leq p \leq n}$. Plus précisément, si on remplace dans (1.3) le couple X_2 , et X_1 , par X_n , et (X_0, \dots, X_{n-1}) , alors on obtient la formule récursive suivante

$$\begin{aligned} & \mathbb{P}^{(X_1, \dots, X_n) | X_0}(x_1, \dots, x_n | x_0) \\ &= \mathbb{P}^{X_n | (X_0, \dots, X_{n-1})}(x_n | x_0, \dots, x_{n-1}) \mathbb{P}^{(X_1, \dots, X_{n-1}) | X_0}(x_1, \dots, x_{n-1} | x_0) \end{aligned}$$

On en déduit aisément la formule multiplicative

$$\mathbb{P}^{(X_1, \dots, X_n) | X_0}(x_1, \dots, x_n | x_0) = \left[\prod_{p=1}^n \mathbb{P}^{X_p | (X_0, \dots, X_{p-1})}(x_p | x_0, \dots, x_{p-1}) \right]$$

3. **[Formule de décorrélation.]** Pour tout couple de fonctions $f_i : E_i \rightarrow \mathbb{R}$, avec $i = 1, 2$, on a

$$\begin{aligned} & \mathbb{E}(f_1(X_1) f_2(X_2)) \\ &= \sum_{(x_1, x_2) \in (E_1 \times E_2)} f_1(x_1) f_2(x_2) \mathbb{P}^{X_2 | X_1}(x_2 | x_1) \mathbb{P}^{X_1}(x_1) \\ &= \sum_{x_1 \in E_1} \left[\sum_{x_2 \in E_2} f_2(x_2) \mathbb{P}^{X_2 | X_1}(x_2 | x_1) \right] f_1(x_1) \mathbb{P}^{X_1}(x_1) \\ &= \mathbb{E}(\mathbb{E}(f_2(X_2) | X_1) f_1(X_1)) \end{aligned}$$

Par conséquent, pour toute fonction f_2 nous avons la propriété suivante

$$\forall Y_1 \subseteq \sigma(X_1) \quad \mathbb{E}(Y_1 f_2(X_2)) = \mathbb{E}(Y_1 \mathbb{E}(f_2(X_2) | X_1)) \quad (1.5)$$

4. **[Indépendance et conditionnement].** Lorsque les variables X_1 et X_2 sont indépendantes, on remarque que

$$\mathbb{P}^{(X_1, X_2)}(x_1, x_2) = \mathbb{P}^{X_1}(x_1) \mathbb{P}^{X_2}(x_2) \Rightarrow \begin{cases} \mathbb{P}^{X_2 | X_1}(x_2 | x_1) = \mathbb{P}^{X_2}(x_2) \\ \mathbb{P}^{X_1 | X_2}(x_1 | x_2) = \mathbb{P}^{X_1}(x_1) \end{cases}$$

Exercice 1.3.5 Soit $(X_p)_{0 \leq p \leq n}$ une suite de v.a. définies sur un même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans des espaces au plus dénombrables $(E_p, \mathcal{P}(E_p))_{0 \leq p \leq n}$. Supposons que la loi conditionnelle de la séquence (X_1, \dots, X_n) en l'évènement $\{X_0 = x_0\}$ soit de la forme

$$\begin{aligned} & \mathbb{P}^{(X_1, \dots, X_n) | X_0}(x_1, \dots, x_n | x_0) \\ &= p_n(x_n | x_0, \dots, x_{n-1}) p_{n-1}(x_{n-1} | x_0, \dots, x_{n-2}) \dots p_2(x_2 | x_0, x_1) p_1(x_1 | x_0) \end{aligned} \quad (1.6)$$

Dans la formule précédente, p_k désigne une suite de fonctions boréliennes positives, et telles que $\sum_{x_k} p_k(x_k | x_0, \dots, x_{k-1}) = 1$, pour tout $k \geq 1$, et pour toute séquence $(x_0, \dots, x_{k-1}) \in (E_0 \times \dots \times E_{k-1})$. Montrer que l'on a nécessairement

$$\mathbb{P}^{X_k | (X_0, \dots, X_{k-1})}(x_k | x_0, \dots, x_{k-1}) = p_k(x_k | x_0, \dots, x_{k-1})$$

Exercice 1.3.6 Soit (X_1, X_2) un couple de v.a. à valeurs dans un espace produit fini $(E_1 \times E_2)$, muni de l'algèbre discrète $\mathcal{P}(E_1 \times E_2)$. Montrer que pour toute fonction $f : E_2 \rightarrow \mathbb{R}$, on a la formule

$$\begin{aligned} & \mathbb{E}((f(X_2) - \mathbb{E}(f(X_2) | X_1))^2) \\ &= \inf \{ \mathbb{E}((f(X_2) - Y_1)^2) : Y_1 \text{ v.a. } \sigma(X_1)\text{-mesurable} \} \end{aligned}$$

Cette description variationnelle montre l'espérance conditionnelle est un estimateur optimal, en ce sens où elle minimise la variance d'erreur.

Exercice 1.3.7 Soit (X_1, X_2, X_3) un triplé de v.a. à valeurs dans un espace produit fini $(E_1 \times E_2 \times E_3)$, muni de l'algèbre discrète. Montrer que pour toute fonction $f : E_3 \rightarrow \mathbb{R}$, on a la formule de conditionnements emboîtés

$$\mathbb{E}(\mathbb{E}(f(X_3) | X_2, X_1) | X_1) = \mathbb{E}(f(X_3) | X_1)$$

